An introduction to variational image processing

Benjamin Berkels

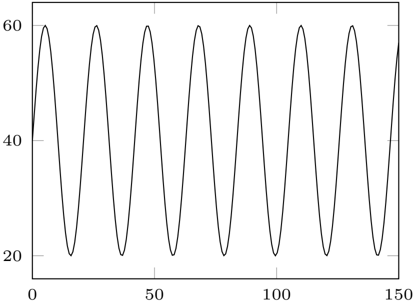
Aachen Institute for Advanced Study in Computational Engineering Science

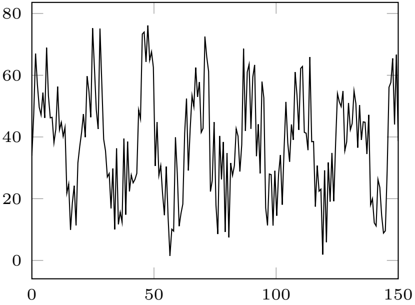
RWTH AACHEN UNIVERSITY

IPAM Long Program on Computational Microscopy

Institute for Pure and Applied Mathematics, UCLA, September 16th, 2022

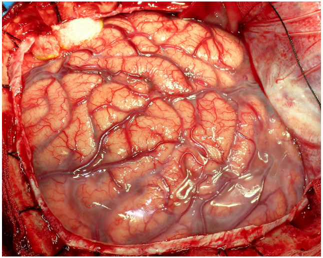
Image modalities





signals

Image modalities

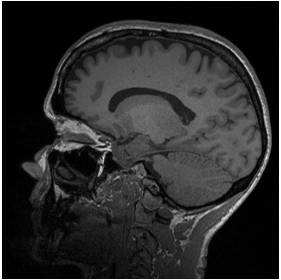
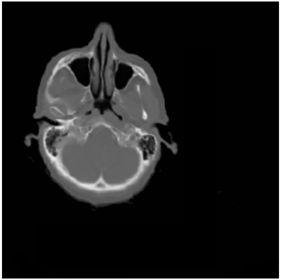


digital photographs

Image modalities

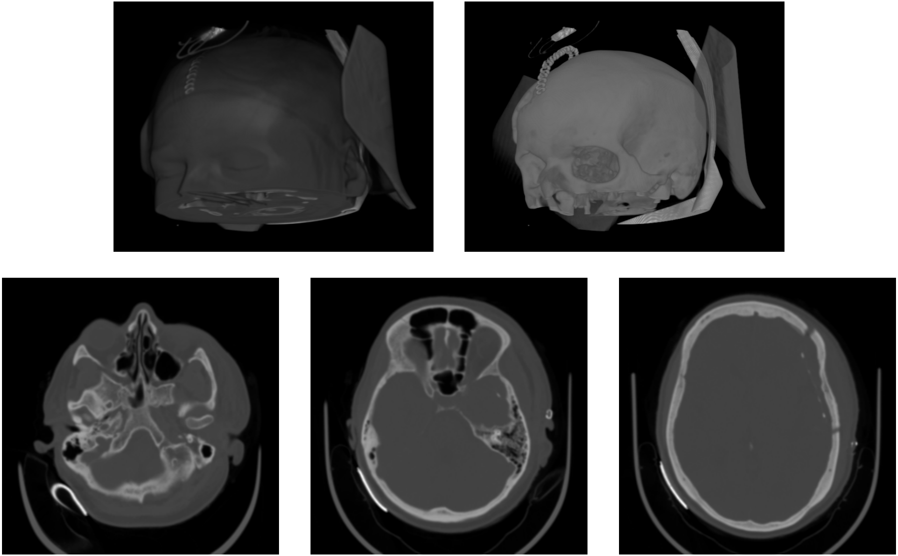
videos

Image modalities

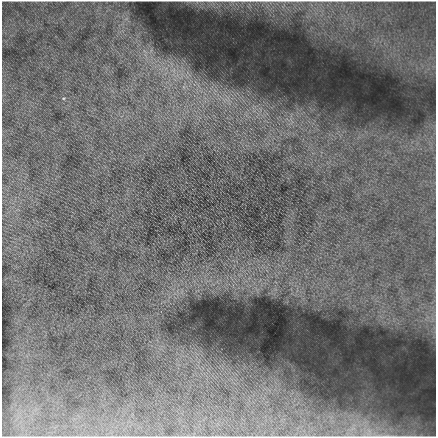

medical modalities (MR / CT / PET)

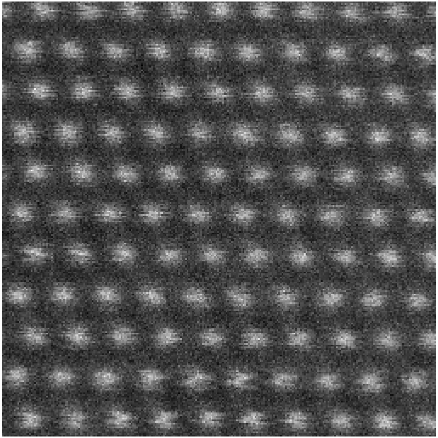
Image modalities



volumetric images

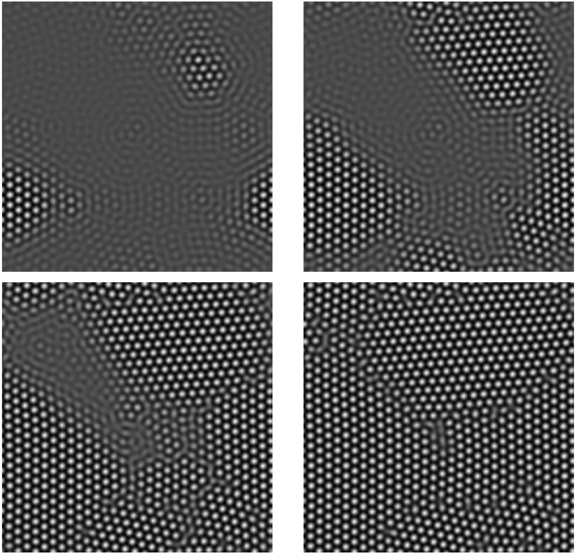
Image modalities





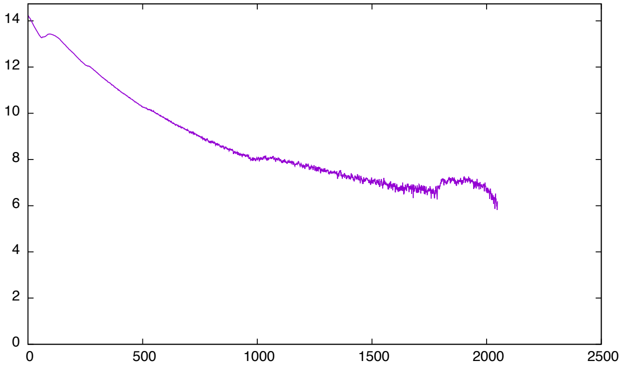
electron microscopy (TEM / STEM)

Image modalities



simulated / computer generated images

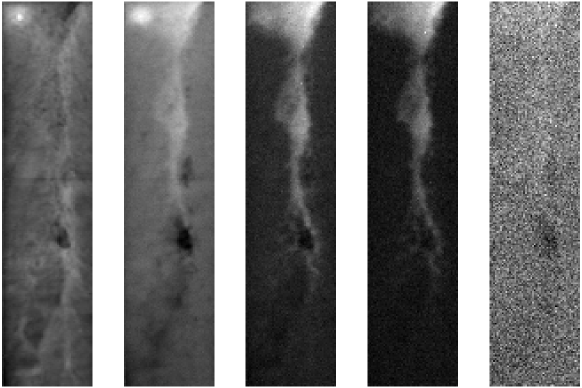
Image modalities



typical EELS spectrum (log scale)

Electron energy loss spectroscopy (EELS)

Image modalities



Electron energy loss spectroscopy (EELS)

Fundamental image processing tasks

Image denoising

Given: Noisy image $f = f_0 + n$.

Task: Recover f_0.

Image deblurring

Given: A blurred $f = Af_0$.

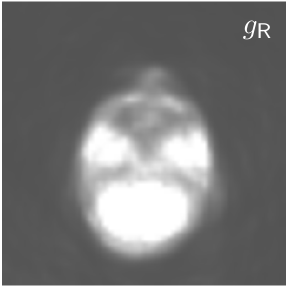
Task: Recover f_0.

Given: Image f showing an object.

Task: Recover the object (as region).

Image registration

Given: Two images g_{T} (template) and g_{R} (reference).



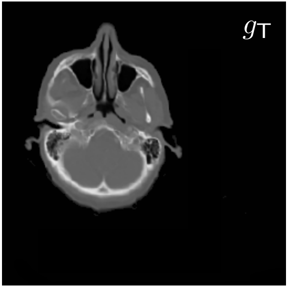
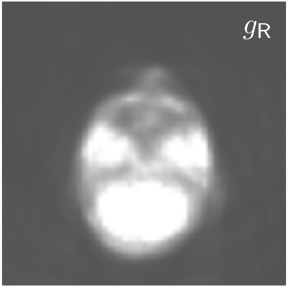
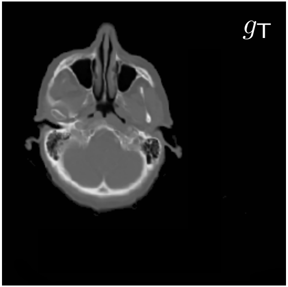


Image registration

Given: Two images g_T (template) and g_R (reference).





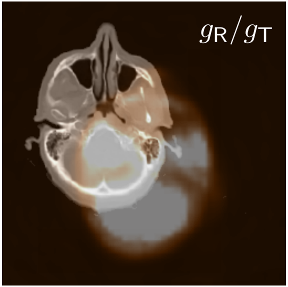
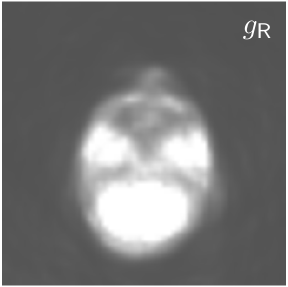
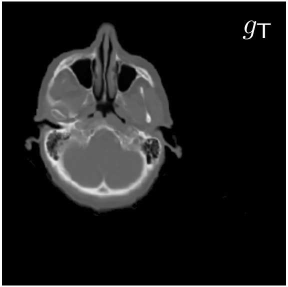


Image registration

Given: Two images g_T (template) and g_R (reference).





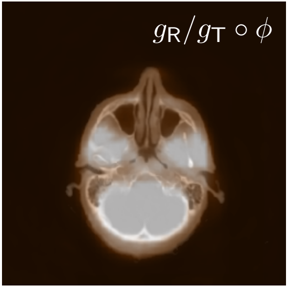


Image registration

Given: Two images g_{T} (template) and g_{R} (reference).

Task: Find a deformation $\phi : \Omega \to \Omega$ (here $\Omega = [0,1]^2$) such that

$$g_{\mathsf{T}} \circ \phi \sim g_{\mathsf{R}}.$$

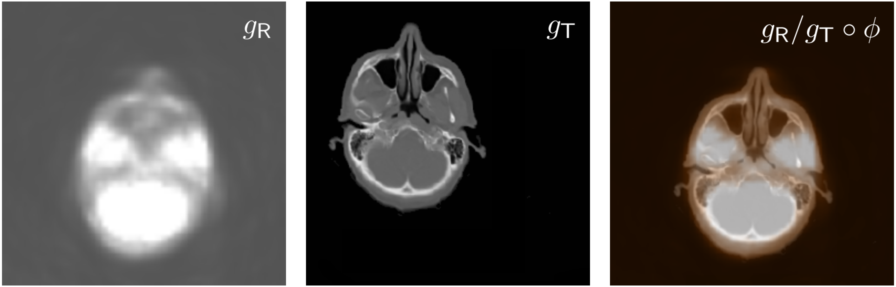
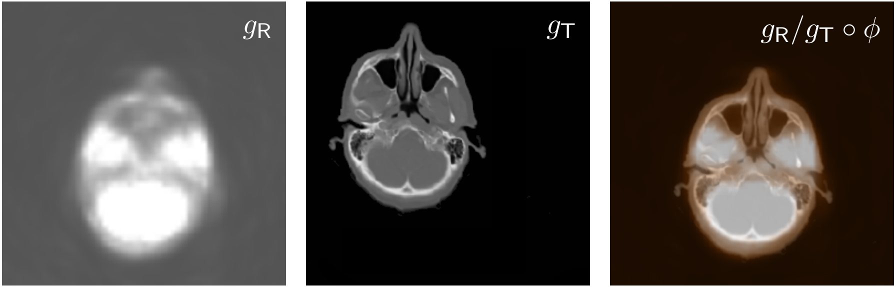


Image registration

Given: Two images g_T (template) and g_R (reference).

Task: Find a deformation $\phi : \Omega \to \Omega$ (here $\Omega = [0,1]^2$) such that

$$g_\mathsf{T} \circ \phi \sim g_\mathsf{R}.$$



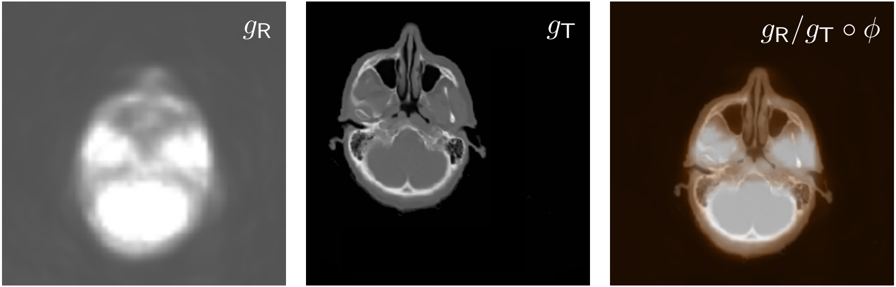
- What is a suitable notion of "$\sim$"?

Image registration

Given: Two images g_T (template) and g_R (reference).

Task: Find a deformation $\phi : \Omega \to \Omega$ (here $\Omega = [0,1]^2$) such that

$$g_\mathsf{T} \circ \phi \sim g_\mathsf{R}.$$



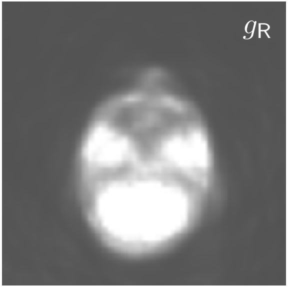
- What is a suitable notion of "$\sim$"? $\Rightarrow$ *Similarity measure*

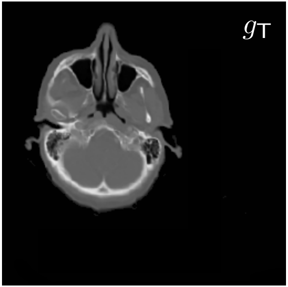
Image registration

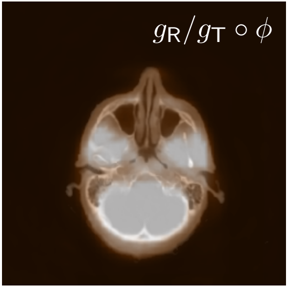
Given: Two images g_T (template) and g_R (reference).

Task: Find a deformation $\phi : \Omega \to \Omega$ (here $\Omega = [0,1]^2$) such that

$$g_\mathsf{T} \circ \phi \sim g_\mathsf{R}.$$







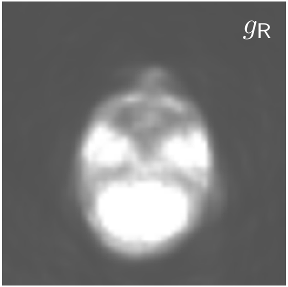
- What is a suitable notion of "$\sim$"? $\Rightarrow$ *Similarity measure*
- How to represent ϕ?

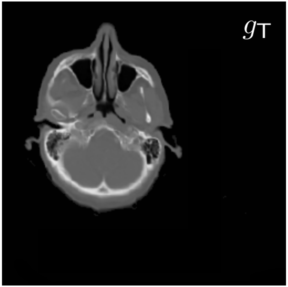
Image registration

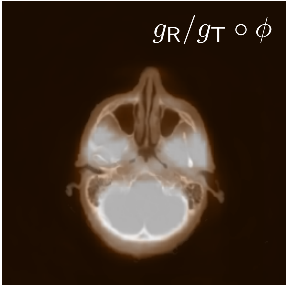
Given: Two images g_T (template) and g_R (reference).

Task: Find a deformation $\phi : \Omega \to \Omega$ (here $\Omega = [0,1]^2$) such that

$$g_\mathsf{T} \circ \phi \sim g_\mathsf{R}.$$







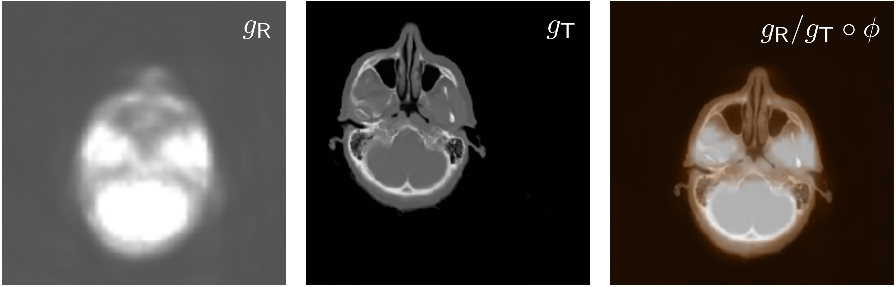
- What is a suitable notion of "$\sim$"? $\Rightarrow$ *Similarity measure*
- How to represent ϕ? $\Rightarrow$ *Regularization or parametrization*

Image registration

Given: Two images g_T (template) and g_R (reference).

Task: Find a deformation $\phi : \Omega \to \Omega$ (here $\Omega = [0,1]^2$) such that

$$g_\mathsf{T} \circ \phi \sim g_\mathsf{R}.$$



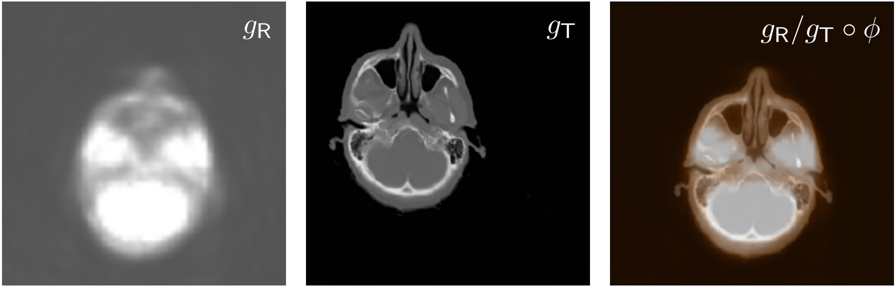
- What is a suitable notion of "$\sim$"? $\Rightarrow$ *Similarity measure*
- How to represent ϕ? $\Rightarrow$ *Regularization or parametrization*
- How to compute the deformation ϕ numerically?

Image registration

Given: Two images g_T (template) and g_R (reference).

Task: Find a deformation $\phi : \Omega \to \Omega$ (here $\Omega = [0,1]^2$) such that

$$g_\mathsf{T} \circ \phi \sim g_\mathsf{R}.$$



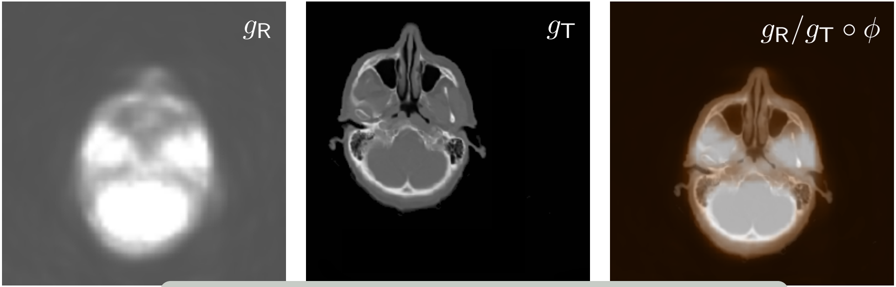
- What is a suitable notion of "$\sim$"? $\Rightarrow$ *Similarity measure*
- How to represent ϕ? $\Rightarrow$ *Regularization or parametrization*
- How to compute the deformation ϕ numerically? $\Rightarrow$ *Optimization*

Image registration

Given: Two images g_{T} (template) and g_{R} (reference).

Task: Find a deformation $\phi : \Omega \to \Omega$ (here $\Omega = [0,1]^2$) such that

$$g_{\mathsf{T}} \circ \phi \sim g_{\mathsf{R}}.$$



- What is a
- How to r
- How to compute the deformation ϕ numerically? $\Rightarrow$ *Optimization*

Introductory text books on image registration:
[Modersitzki, '04][Modersitzki, '09]

Idea to approach a wide range of problems

Variational approaches - Image registration

Idea to approach a wide range of problems

- rephrase task as conditions on the solution

Variational approaches - Image registration

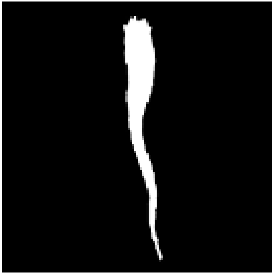
Idea to approach a wide range of problems

- rephrase task as conditions on the solution
- find a function measuring how well the conditions are fulfilled

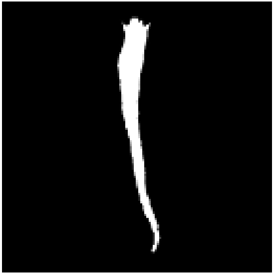
Idea to approach a wide range of problems

- rephrase task as conditions on the solution
- find a function measuring how well the conditions are fulfilled

Example: Register two images $g_{\mathsf{T}}, g_{\mathsf{R}}$ with comparable gray values.



g_{R}



g_{T}

Variational approaches - Image registration

Idea to approach a wide range of problems

- rephrase task as conditions on the solution
- find a function measuring how well the conditions are fulfilled

Example: Register two images $g_\mathsf{T}, g_\mathsf{R}$ with comparable gray values.

Conditions on the non-parametric, non-rigid deformation ϕ:

Variational approaches - Image registration

Idea to approach a wide range of problems

- rephrase task as conditions on the solution
- find a function measuring how well the conditions are fulfilled

Example: Register two images $g_{\mathsf{T}}, g_{\mathsf{R}}$ with comparable gray values.

Conditions on the non-parametric, non-rigid deformation ϕ:

- $g_{\mathsf{T}} \circ \phi \sim g_{\mathsf{R}}$

Variational approaches - Image registration

Idea to approach a wide range of problems

- rephrase task as conditions on the solution
- find a function measuring how well the conditions are fulfilled

Example: Register two images $g_\mathsf{T}, g_\mathsf{R}$ with comparable gray values.

Conditions on the non-parametric, non-rigid deformation ϕ:

- $g_\mathsf{T} \circ \phi \sim g_\mathsf{R} \;\Rightarrow g_\mathsf{T}(\phi(x)) \approx g_\mathsf{R}(x)$ for all pixels x

Variational approaches - Image registration

Idea to approach a wide range of problems

- rephrase task as conditions on the solution
- find a function measuring how well the conditions are fulfilled

Example: Register two images $g_\mathsf{T}, g_\mathsf{R}$ with comparable gray values.

Conditions on the non-parametric, non-rigid deformation ϕ:

- $g_\mathsf{T} \circ \phi \sim g_\mathsf{R} \;\Rightarrow g_\mathsf{T}(\phi(x)) \approx g_\mathsf{R}(x)$ for all pixels x
- ϕ should be smooth

Idea to approach a wide range of problems

- rephrase task as conditions on the solution
- find a function measuring how well the conditions are fulfilled

Example: Register two images $g_\mathsf{T}, g_\mathsf{R}$ with comparable gray values.

Conditions on the non-parametric, non-rigid deformation ϕ:

- $g_\mathsf{T} \circ \phi \sim g_\mathsf{R} \;\Rightarrow g_\mathsf{T}(\phi(x)) \approx g_\mathsf{R}(x)$ for all pixels x
- ϕ should be smooth $\Rightarrow D(\phi - \mathrm{id})$ is mostly small

Variational approaches - Image registration

Idea to approach a wide range of problems

- rephrase task as conditions on the solution
- find a function measuring how well the conditions are fulfilled

Example: Register two images $g_\mathsf{T}, g_\mathsf{R}$ with comparable gray values.

Conditions on the non-parametric, non-rigid deformation ϕ:

- $g_\mathsf{T} \circ \phi \sim g_\mathsf{R} \;\Rightarrow g_\mathsf{T}(\phi(x)) \approx g_\mathsf{R}(x)$ for all pixels x
- ϕ should be smooth $\Rightarrow D(\phi - \mathrm{id})$ is mostly small

$$E[\phi] = \frac{1}{2} \int_\Omega |g_\mathsf{T}(\phi(x)) - g_\mathsf{R}(x)|^2 \,\mathrm{d}x$$

Idea to approach a wide range of problems

- rephrase task as conditions on the solution
- find a function measuring how well the conditions are fulfilled

Example: Register two images $g_\mathsf{T}, g_\mathsf{R}$ with comparable gray values.

Conditions on the non-parametric, non-rigid deformation ϕ:

- $g_\mathsf{T} \circ \phi \sim g_\mathsf{R} \;\Rightarrow g_\mathsf{T}(\phi(x)) \approx g_\mathsf{R}(x)$ for all pixels x
- ϕ should be smooth $\Rightarrow D(\phi - \mathrm{id})$ is mostly small

$$E[\phi] = \frac{1}{2}\int_\Omega |g_\mathsf{T}(\phi(x)) - g_\mathsf{R}(x)|^2 \,\mathrm{d}x + \frac{\lambda}{2}\int_\Omega \|D\phi(x) - \mathbb{1}\|^2 \,\mathrm{d}x$$

Idea to approach a wide range of problems

- rephrase task as conditions on the solution
- find a function measuring how well the conditions are fulfilled

Example: Register two images $g_{\mathsf{T}}, g_{\mathsf{R}}$ with comparable gray values.

Conditions on the non-parametric, non-rigid deformation ϕ:

- $g_{\mathsf{T}} \circ \phi \sim g_{\mathsf{R}} \;\Rightarrow g_{\mathsf{T}}(\phi(x)) \approx g_{\mathsf{R}}(x)$ for all pixels x
- ϕ should be smooth $\Rightarrow D(\phi - \mathrm{id})$ is mostly small

$$\begin{aligned} E[\phi] &= \frac{1}{2}\int_\Omega |g_{\mathsf{T}}(\phi(x)) - g_{\mathsf{R}}(x)|^2 \,\mathrm{d}x + \frac{\lambda}{2}\int_\Omega \|D\phi(x) - \mathbb{1}\|^2 \,\mathrm{d}x \\ &= \tfrac{1}{2}\|\mathcal{A}[\phi] - g_{\mathsf{R}}\|^2_{L^2} + \tfrac{\lambda}{2}\|D\phi - \mathbb{1}\|^2_{L^2} \end{aligned}$$

Variational approaches - Image registration

Idea to approach a wide range of problems

- rephrase task as conditions on the solution
- find a function measuring how well the conditions are fulfilled

Example: Register two images $g_\mathsf{T}, g_\mathsf{R}$ with comparable gray values.

Conditions on the non-parametric, non-rigid deformation ϕ:

- $g_\mathsf{T} \circ \phi \sim g_\mathsf{R} \;\Rightarrow g_\mathsf{T}(\phi(x)) \approx g_\mathsf{R}(x)$ for all pixels x
- ϕ should be smooth $\Rightarrow D(\phi - \mathrm{id})$ is mostly small

$$
\begin{aligned}
E[\phi] &= \frac{1}{2}\int_\Omega \left|g_\mathsf{T}(\phi(x)) - g_\mathsf{R}(x)\right|^2 \mathrm{d}x + \frac{\lambda}{2}\int_\Omega \left\|D\phi(x) - \mathbb{1}\right\|^2 \mathrm{d}x \\
&= \tfrac{1}{2}\left\|\mathcal{A}[\phi] - g_\mathsf{R}\right\|_{L^2}^2 + \tfrac{\lambda}{2}\left\|D\phi - \mathbb{1}\right\|_{L^2}^2
\end{aligned}
$$

$\Rightarrow g_\mathsf{T}$ and g_R are registered by minimizing E.

Variational approaches - Image registration

Idea to approach a wide range of problems

- rephrase task as conditions on the solution
- find a function measuring how well the conditions are fulfilled

Example: Register two images $g_\mathsf{T}, g_\mathsf{R}$ with comparable gray values.

Conditions on the non-parametric, non-rigid deformation ϕ:

- $g_\mathsf{T} \circ \phi \sim g_\mathsf{R} \;\Rightarrow g_\mathsf{T}(\phi(x)) \approx g_\mathsf{R}(x)$ for all pixels x
- ϕ should be smooth $\Rightarrow D(\phi - \mathrm{id})$ is mostly small

$$\begin{aligned} E[\phi] &= \frac{1}{2}\int_\Omega |g_\mathsf{T}(\phi(x)) - g_\mathsf{R}(x)|^2 \, \mathrm{d}x + \frac{\lambda}{2}\int_\Omega \|D\phi(x) - \mathbb{1}\|^2 \, \mathrm{d}x \\ &= \tfrac{1}{2}\,\|\mathcal{A}[\phi] - g_\mathsf{R}\|_{L^2}^2 + \tfrac{\lambda}{2}\,\|D\phi - \mathbb{1}\|_{L^2}^2 \end{aligned}$$

$\Rightarrow$ g_T and g_R are registered by minimizing E.

Structure "$\mathcal{D}[\phi] + \lambda\mathcal{R}[\phi]$" typical for registration (and inverse problems).

Idea to approach a wide range of problems

- rephrase task as conditions on the solution
- find a function measuring how well the conditions are fulfilled

Example: Register two images $g_\mathsf{T}, g_\mathsf{R}$ with comparable gray values.

Conditions on the non-parametric, non-rigid deformation ϕ:

- $g_\mathsf{T} \circ \phi \sim g_\mathsf{R} \;\Rightarrow g_\mathsf{T}(\phi(x)) \approx g_\mathsf{R}(x)$ for all pixels x
- ϕ should be smooth $\Rightarrow D(\phi - \mathrm{id})$ is mostly small

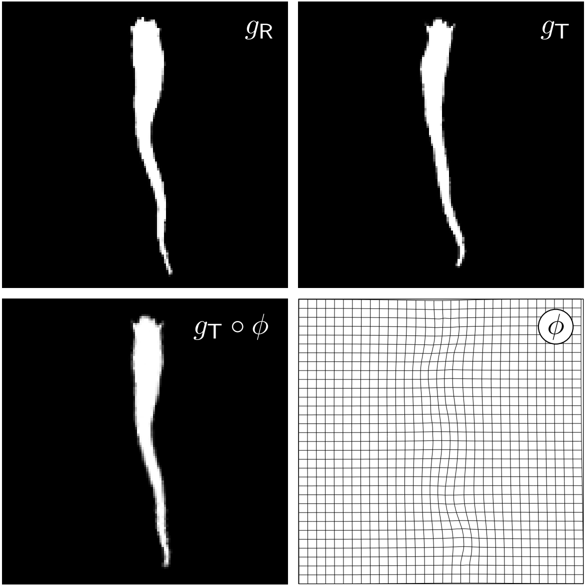
$$
\begin{aligned}
E[\phi] &= \frac{1}{2} \int_\Omega |g_\mathsf{T}(\phi(x)) - g_\mathsf{R}(x)|^2 \,\mathrm{d}x + \frac{\lambda}{2} \int_\Omega \|D\phi(x) - \mathbb{1}\|^2 \,\mathrm{d}x \\
&= \tfrac{1}{2} \|\mathcal{A}[\phi] - g_\mathsf{R}\|_{L^2}^2 + \tfrac{\lambda}{2} \|D\phi - \mathbb{1}\|_{L^2}^2
\end{aligned}
$$

$\Rightarrow g_\mathsf{T}$ and g_R are registered by minimizing E.

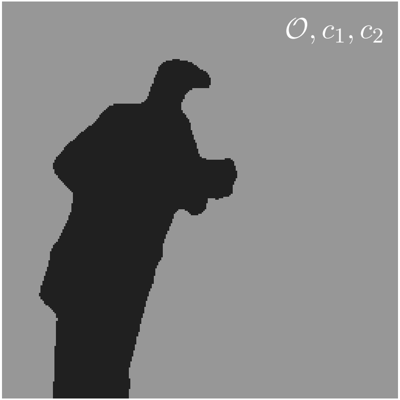
Structure "$\mathcal{D}[\phi] + \lambda\mathcal{R}[\phi]$" typical for registration (and inverse problems).

Note: Competing conditions are balanced by weights (like λ in E)

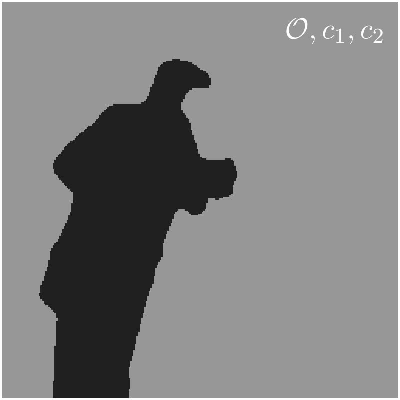
A simple registration result




g



Given an image g, we search for a piecewise constant segmentation, i.e. two gray values c_1, c_2 and a region $\mathcal{O}$.



Given an image g, we search for a piecewise constant segmentation, i.e. two gray values c_1, c_2 and a region $\mathcal{O}$. A valid segmentation minimizes

$$E[\mathcal{O}, c_1, c_2] = \int_{\mathcal{O}} (g(x) - c_1)^2 \, \mathrm{d}x + \int_{\Omega \setminus \mathcal{O}} (g(x) - c_2)^2 \, \mathrm{d}x + \nu \, \mathrm{Per}(\mathcal{O}).$$

(Piecewise constant binary Mumford-Shah model)

Denoising

Given a noise image f, i.e. $f = f_0 + n$, find f_0 by minimizing

$$J[u] = \underbrace{\int_\Omega (u - f)^2 \,\mathrm{d}x}_{\text{data term}} + \lambda \underbrace{\int_\Omega \|\nabla u(x)\| \,\mathrm{d}x}_{\text{regularizer}} \quad \text{(Rudin-Osher-Fatemi)}$$

Variational approaches - Denoising and deblurring

Denoising

Given a noise image f, i.e. $f = f_0 + n$, find f_0 by minimizing

$$J[u] = \underbrace{\int_\Omega (u - f)^2 \,\mathrm{d}x}_{\text{data term}} + \lambda \underbrace{\int_\Omega \|\nabla u(x)\| \,\mathrm{d}x}_{\text{regularizer}} \quad \text{(Rudin-Osher-Fatemi)}$$

Deconvolution/Deblurring

Given a blurry image/signal f, i.e. $f = Af_0$, find f_0 by minimizing

$$J[u] = \int_\Omega (Au - f)^2 \,\mathrm{d}x + \lambda \int_\Omega \|\nabla u(x)\| \,\mathrm{d}x\,.$$

General task

Given: Normed vector space $(X, \|\cdot\|), M \subset X, J : M \to \mathbb{R}$,

Find: $y^* \in M$ such that $J[y^*] \leq J[y]$ for all $y \in M$.

General task

Given: Normed vector space $(X, \|\cdot\|), M \subset X, J : M \to \mathbb{R}$,

Find: $y^* \in M$ such that $J[y^*] \leq J[y]$ for all $y \in M$.

Here,

- J is called *objective functional*,
- M is the *admissible set*.

General task

Given: Normed vector space $(X, \|\cdot\|), M \subset X, J : M \to \mathbb{R}$,

Find: $y^* \in M$ such that $J[y^*] \leq J[y]$ for all $y \in M$.

Here,

- J is called *objective functional*,
- M is the *admissible set*.

Note: In the following, vector space always means real vector space.

General task

Given: Normed vector space $(X, \|\cdot\|), M \subset X, J : M \to \mathbb{R}$,

Find: $y^* \in M$ such that $J[y^*] \leq J[y]$ for all $y \in M$.

Here,

- J is called *objective functional*,
- M is the *admissible set*.

Note: In the following, vector space always means real vector space.

- structure is very similar to classical optimization,
- but $\dim(X) = \infty$ is possible, e.g. if X is a function space.

General task

Given: Normed vector space $(X, \|\cdot\|), M \subset X, J : M \to \mathbb{R}$,

Find: $y^* \in M$ such that $J[y^*] \leq J[y]$ for all $y \in M$.

Here,

- J is called *objective functional*,
- M is the *admissible set*.

Note: In the following, vector space always means real vector space.

- structure is very similar to classical optimization,
- but $\dim(X) = \infty$ is possible, e.g. if X is a function space.

Central questions

- Existence of minimizers? $\dim(X) = \infty$ has large implications.
- Characterization of minimizers? (necessary/sufficient conditions)
- How can they be efficiently computed in practice?

Variational image registration

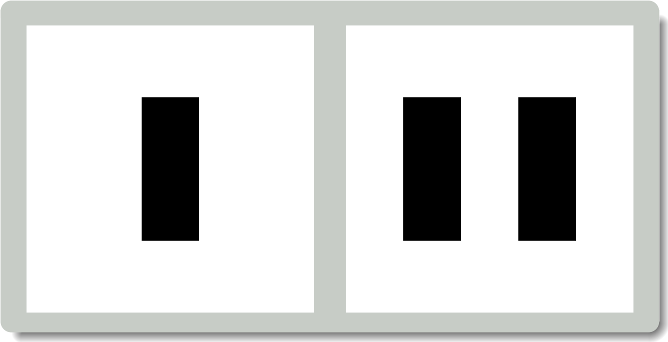
Non-parametric deformations

Minimization with respect to a deformation ϕ is highly ill-posed:

Non-parametric deformations

Minimization with respect to a deformation ϕ is highly ill-posed:

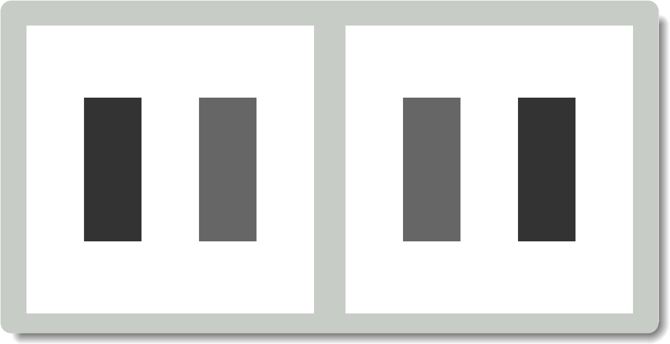
- solution not unique



Non-parametric deformations

Minimization with respect to a deformation ϕ is highly ill-posed:

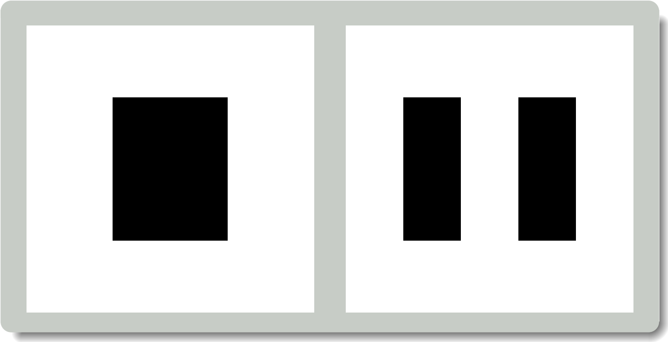
- solution not unique
- small input changes can lead to large output changes



Non-parametric deformations

Minimization with respect to a deformation ϕ is highly ill-posed:

- solution not unique
- small input changes can lead to large output changes
- obtained deformation could be discontinuous



Non-parametric deformations

Minimization with respect to a deformation ϕ is highly ill-posed:

- solution not unique
- small input changes can lead to large output changes
- obtained deformation could be discontinuous

Only some of these problems are resolved by the regularization.

Non-parametric deformations

Minimization with respect to a deformation ϕ is highly ill-posed:

- solution not unique
- small input changes can lead to large output changes
- obtained deformation could be discontinuous

Only some of these problems are resolved by the regularization.

For instance, proper regularization guarantees existence of minimizers.

Non-parametric deformations

Minimization with respect to a deformation ϕ is highly ill-posed:

- solution not unique
- small input changes can lead to large output changes
- obtained deformation could be discontinuous

Only some of these problems are resolved by the regularization.

For instance, proper regularization guarantees existence of minimizers.

Still, a sophisticated minimization approach is necessary:

Non-parametric deformations

Minimization with respect to a deformation ϕ is highly ill-posed:

- solution not unique
- small input changes can lead to large output changes
- obtained deformation could be discontinuous

Only some of these problems are resolved by the regularization.

For instance, proper regularization guarantees existence of minimizers.

Still, a sophisticated minimization approach is necessary:

- multilevel strategy

Non-parametric deformations

Minimization with respect to a deformation ϕ is highly ill-posed:

- solution not unique
- small input changes can lead to large output changes
- obtained deformation could be discontinuous

Only some of these problems are resolved by the regularization.

For instance, proper regularization guarantees existence of minimizers.

Still, a sophisticated minimization approach is necessary:

- multilevel strategy
- regularized gradient descent (gradient flow)

Non-parametric deformations

Minimization with respect to a deformation ϕ is highly ill-posed:

- solution not unique
- small input changes can lead to large output changes
- obtained deformation could be discontinuous

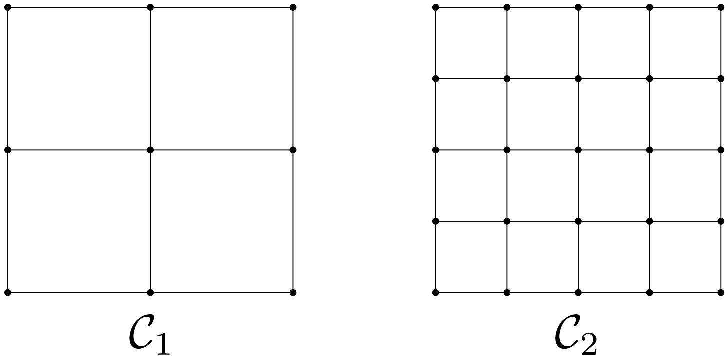
Only some of these problems are resolved by the regularization.

For instance, proper regularization guarantees existence of minimizers.

Still, a sophisticated minimization approach is necessary:

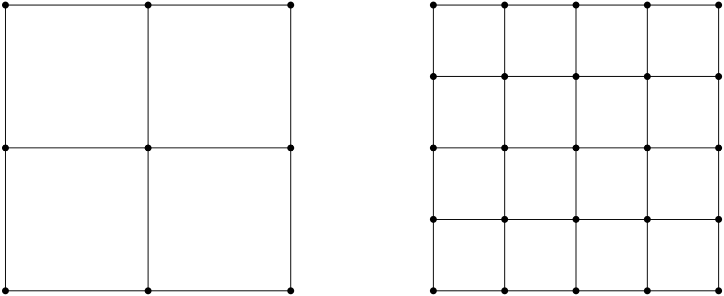
- multilevel strategy
- regularized gradient descent (gradient flow)
- step size control (e. g. Armijo rule)

Multilevel minimization



- hierarchy of nested meshes $\mathcal{C}_1 \subset \mathcal{C}_2 \subset \mathcal{C}_3 \subset \ldots$

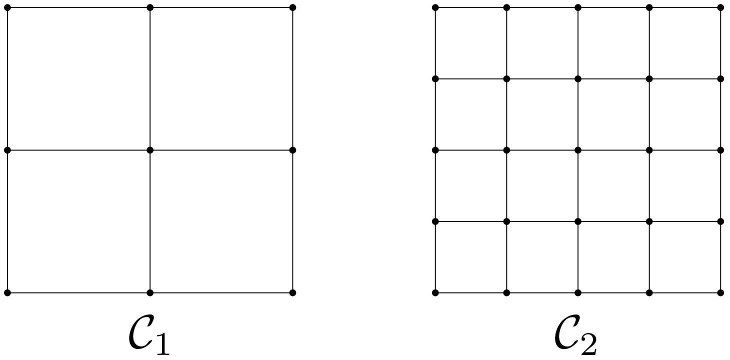
Multilevel minimization



$\mathcal{C}_1$ $\mathcal{C}_2$

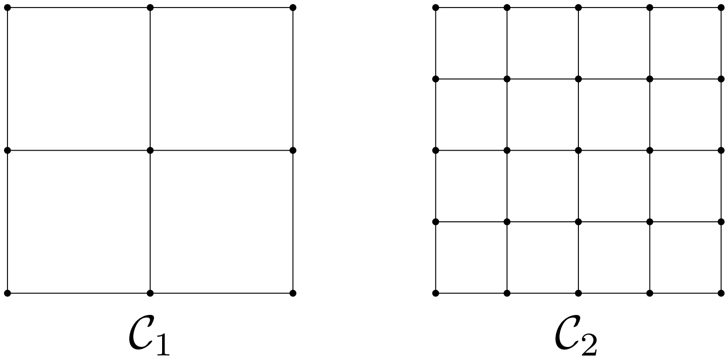
- hierarchy of nested meshes $\mathcal{C}_1 \subset \mathcal{C}_2 \subset \mathcal{C}_3 \subset \ldots$
- solve the problem on a coarse starting level l

Multilevel minimization



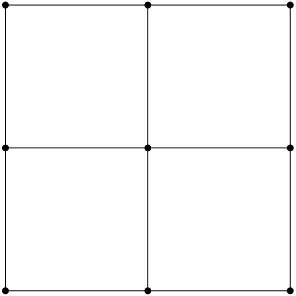
- hierarchy of nested meshes $\mathcal{C}_1 \subset \mathcal{C}_2 \subset \mathcal{C}_3 \subset \ldots$
- solve the problem on a coarse starting level l
- prolongate the result to level $l+1$

Multilevel minimization

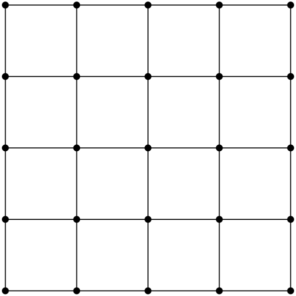


- hierarchy of nested meshes $\mathcal{C}_1 \subset \mathcal{C}_2 \subset \mathcal{C}_3 \subset \dots$
- solve the problem on a coarse starting level l
- prolongate the result to level $l+1$
- solve on level $l+1$ with the prolongated result as initial guess

Multilevel minimization



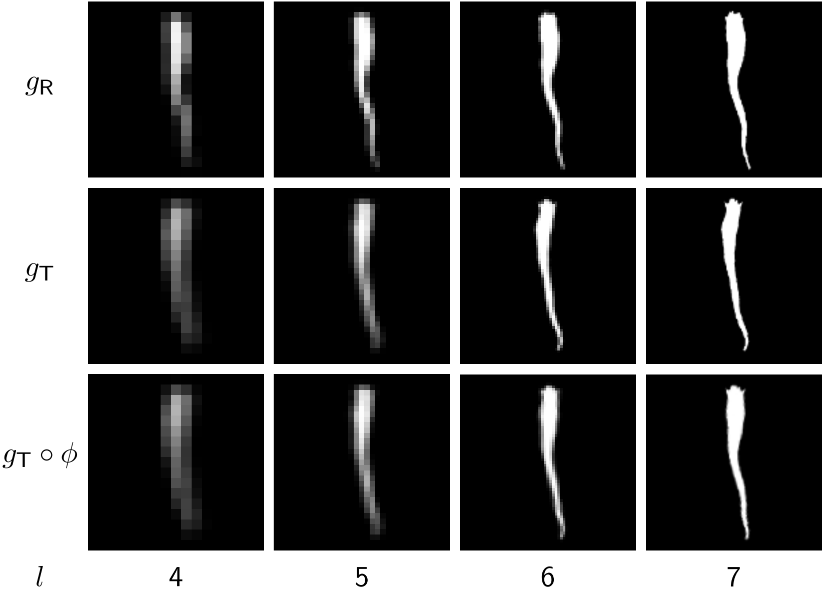
$\mathcal{C}_1$



$\mathcal{C}_2$

- hierarchy of nested meshes $\mathcal{C}_1 \subset \mathcal{C}_2 \subset \mathcal{C}_3 \subset \ldots$
- solve the problem on a coarse starting level l
- prolongate the result to level $l+1$
- solve on level $l+1$ with the prolongated result as initial guess
- iterate till reaching the finest level

Multilevel minimization



Gradient flows

Idea: Minimize E by going in direction of steepest descent:

$$\partial_t \phi = -\operatorname{grad} E[\phi]$$

Gradient flows

Idea: Minimize E by going in direction of steepest descent:

$$\partial_t \phi = -\operatorname{grad} E[\phi]$$

This evolution is attracted by the *nearest* local minimizer.

Gradient flows

Idea: Minimize E by going in direction of steepest descent:

$$\partial_t \phi = -\operatorname{grad} E[\phi]$$

This evolution is attracted by the *nearest* local minimizer.

Recall: The gradient is the best linear approximation, i. e.

$$E[\phi] = E[\phi_0] + \operatorname{grad} E[\phi_0] \cdot (\phi - \phi_0) + o(\|\phi - \phi_0\|).$$

Gradient flows

Idea: Minimize E by going in direction of steepest descent:

$$\partial_t \phi = -\operatorname{grad} E[\phi]$$

This evolution is attracted by the *nearest* local minimizer.

Recall: The gradient is the best linear approximation, i. e.

$$E[\phi] = E[\phi_0] + \underbrace{\operatorname{grad} E[\phi_0] \cdot (\phi - \phi_0)}_{\text{Euclidean inner product}} + o(\|\phi - \phi_0\|).$$

Gradient flows

Idea: Minimize E by going in direction of steepest descent:

$$\partial_t \phi = -\operatorname{grad} E[\phi]$$

This evolution is attracted by the *nearest* local minimizer.

Recall: The gradient is the best linear approximation, i. e.

$$E[\phi] = E[\phi_0] + \underbrace{\operatorname{grad} E[\phi_0] \cdot (\phi - \phi_0)}_{\text{Euclidean inner product}} + o(\|\phi - \phi_0\|).$$

$\Rightarrow$ The gradient descent perceives distance in the Euclidean sense.

Gradient flows

Idea: Minimize E by going in direction of steepest descent:

$$\partial_t \phi = -\operatorname{grad} E[\phi]$$

This evolution is attracted by the *nearest* local minimizer.

Recall: The gradient is the best linear approximation, i. e.

$$E[\phi] = E[\phi_0] + \underbrace{\operatorname{grad} E[\phi_0] \cdot (\phi - \phi_0)}_{\text{Euclidean inner product}} + o(\|\phi - \phi_0\|).$$

$\Rightarrow$ The gradient descent perceives distance in the Euclidean sense.

We can define a gradient for any inner product g via

$$E[\phi] = E[\phi_0] + g(\operatorname{grad}_g E[\phi_0], (\phi - \phi_0)) + o(\|\phi - \phi_0\|_g).$$

Gradient flows

Idea: Minimize E by going in direction of steepest descent:

$$\partial_t \phi = -\operatorname{grad} E[\phi]$$

This evolution is attracted by the *nearest* local minimizer.

Recall: The gradient is the best linear approximation, i. e.

$$E[\phi] = E[\phi_0] + \underbrace{\operatorname{grad} E[\phi_0] \cdot (\phi - \phi_0)}_{\text{Euclidean inner product}} + o(\|\phi - \phi_0\|).$$

$\Rightarrow$ The gradient descent perceives distance in the Euclidean sense.

We can define a gradient for any inner product g via

$$E[\phi] = E[\phi_0] + g(\operatorname{grad}_g E[\phi_0], (\phi - \phi_0)) + o(\|\phi - \phi_0\|_g).$$

$\Rightarrow$ Changing the inner product changes which minimizer is *nearest*.

Gradient flows

Idea: Minimize E by going in direction of steepest descent:

$$\partial_t \phi = -\operatorname{grad} E[\phi]$$

This evolution is attracted by the *nearest* local minimizer.

Recall: The gradient is the best linear approximation, i. e.

$$E[\phi] = E[\phi_0] + \underbrace{\operatorname{grad} E[\phi_0] \cdot (\phi - \phi_0)}_{\text{Euclidean inner product}} + o(\|\phi - \phi_0\|).$$

$\Rightarrow$ The gradient descent perceives distance in the Euclidean sense.

We can define a gradient for any inner product g via

$$E[\phi] = E[\phi_0] + g(\operatorname{grad}_g E[\phi_0], (\phi - \phi_0)) + o(\|\phi - \phi_0\|_g).$$

$\Rightarrow$ Changing the inner product changes which minimizer is *nearest*.

$\Rightarrow$ Use an inner product favoring desired solution properties.

Gradient flows

Idea: Minimize E by going in direction of steepest descent:

$$\partial_t \phi = -\operatorname{grad}_g E[\phi]$$

Gradient flows

Idea: Minimize E by going in direction of steepest descent:

$$\partial_t \phi = -\operatorname{grad}_g E[\phi]$$

Denoting the inner product representation by A, we get

$$\partial_t \phi = -A^{-1} DE[\phi]$$

Gradient flows

Idea: Minimize E by going in direction of steepest descent:

$$\partial_t \phi = -\operatorname{grad}_g E[\phi]$$

Denoting the inner product representation by A, we get

$$\partial_t \phi = -A^{-1} DE[\phi]$$

$$\text{time discrete:} \quad \phi^{k+1} = \phi^k - \tau A^{-1} DE[\phi^k].$$

Gradient flows

Idea: Minimize E by going in direction of steepest descent:

$$\partial_t \phi = -\operatorname{grad}_g E[\phi]$$

Denoting the inner product representation by A, we get

$$\partial_t \phi = -A^{-1} DE[\phi]$$

$$\text{time discrete:} \quad \phi^{k+1} = \phi^k - \tau A^{-1} DE[\phi^k].$$

We choose the inner product to be the discrete counterpart of

$$g(v, w) = \int_\Omega vw \, \mathrm{d}x + \frac{\sigma^2}{2} \int_\Omega \nabla v \cdot \nabla w \, \mathrm{d}x$$

Gradient flows

Idea: Minimize E by going in direction of steepest descent:

$$\partial_t \phi = -\operatorname{grad}_g E[\phi]$$

Denoting the inner product representation by A, we get

$$\partial_t \phi = -A^{-1} DE[\phi]$$

$$\text{time discrete:} \quad \phi^{k+1} = \phi^k - \tau A^{-1} DE[\phi^k].$$

We choose the inner product to be the discrete counterpart of

$$g(v,w) = \int_\Omega vw \,\mathrm{d}x + \frac{\sigma^2}{2} \int_\Omega \nabla v \cdot \nabla w \,\mathrm{d}x$$

$\Rightarrow A^{-1}$ "$=$" one implicit heat equation time step with step size $\frac{\sigma^2}{2}$.

Gradient flows

Idea: Minimize E by going in direction of steepest descent:

$$\partial_t \phi = -\operatorname{grad}_g E[\phi]$$

Denoting the inner product representation by A, we get

$$\partial_t \phi = -A^{-1} DE[\phi]$$

$$\text{time discrete:} \quad \phi^{k+1} = \phi^k - \tau A^{-1} DE[\phi^k].$$

We choose the inner product to be the discrete counterpart of

$$g(v, w) = \int_\Omega vw \,\mathrm{d}x + \frac{\sigma^2}{2} \int_\Omega \nabla v \cdot \nabla w \,\mathrm{d}x$$

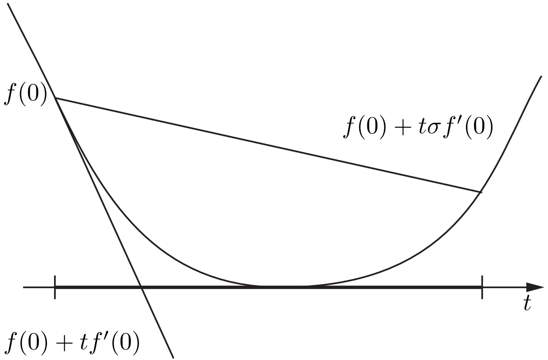
$\Rightarrow A^{-1}$ "$=$" one implicit heat equation time step with step size $\frac{\sigma^2}{2}$.

$\Rightarrow A^{-1}$ is regularizing.

Step size control with the Armijo rule

Finding a suitable step size is a 1D optimization problem:

$$f : \mathbb{R} \to \mathbb{R}, t \mapsto E[\phi + td]$$



Armijo condition

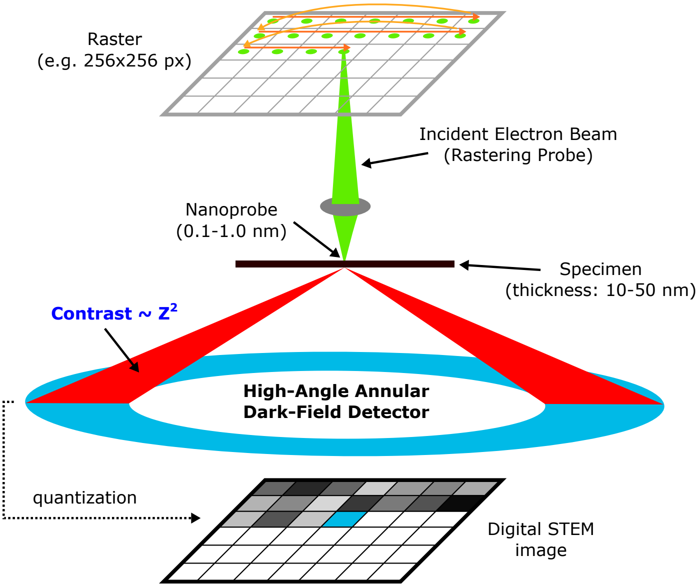
$$\frac{f(\tau) - f(0)}{\tau f'(0)} = \frac{\text{secant slope}}{\text{tangent slope}} \geq \sigma \in (0, 1)$$

High precision analysis of image series from electron microscopy

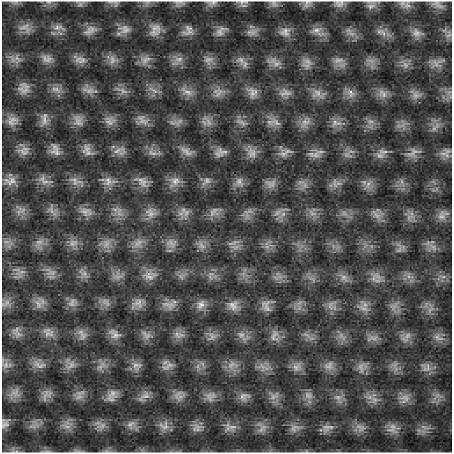
joint work with:

P. Binev, D. A. Blom, R. Sharpley, T. Vogt, W. Dahmen (University of SC), N. Mevenkamp (RWTH, now at Zeiss)
P. Voyles (University of Wisconsin), A. Yankovich (Chalmers)

Scanning transmission electron microscopy

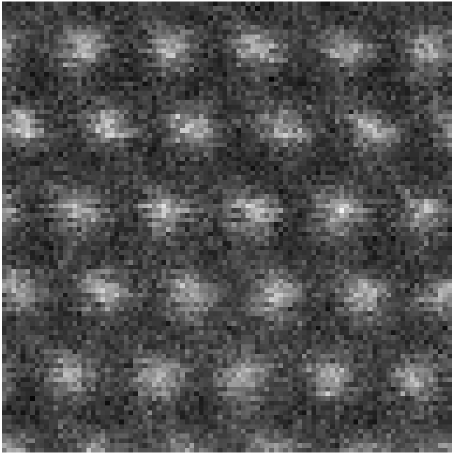


Scanning transmission electron microscopy (cont.)

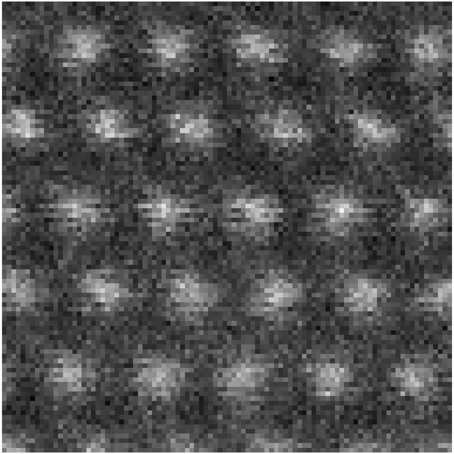


Gallium-Nitrogen lattice at 20.5Mx magnification

Scanning transmission electron microscopy (cont.)



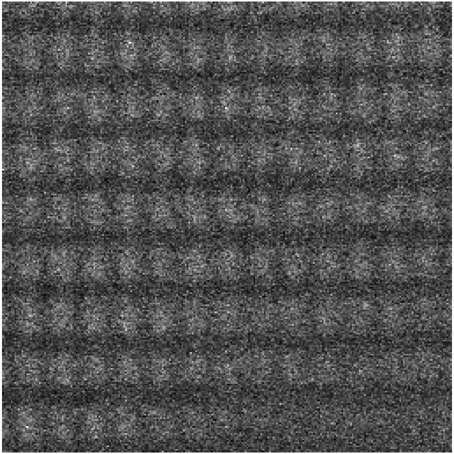
Close-up reveals horizontal distortions and noise



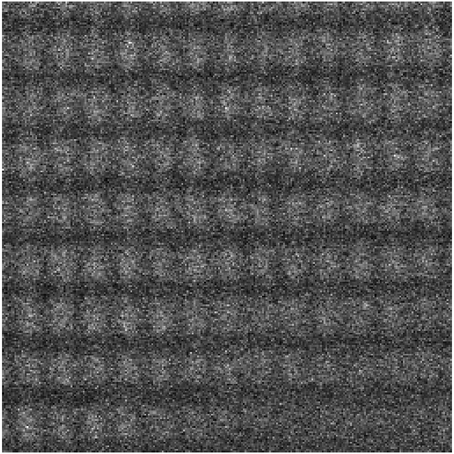
Close-up reveals horizontal distortions and noise

- Distortions arise from environmental and instrumental disturbances

STEM image series of silicon (112)



STEM image series of silicon (112)



Challenges

- individual frames are very noisy
- large movement of the sample during the series acquisition

Normalized cross-correlation

Registration of noisy images requires a robust distance measure.

Normalized cross-correlation

Registration of noisy images requires a robust distance measure.

$$\mathsf{NCC}[f,g] = \frac{1}{|\Omega|}\int_\Omega \frac{\left(f-\overline{f}\right)}{\sigma_f}\frac{(g-\overline{g})}{\sigma_g}\,\mathrm{d}x,$$

where

- $\overline{f} = \frac{1}{|\Omega|}\int_\Omega f\,\mathrm{d}x$ (mean value)
- $\sigma_f = \sqrt{\frac{1}{|\Omega|}\int_\Omega \left(f-\overline{f}\right)^2\mathrm{d}x}$. (standard deviation)

Normalized cross-correlation

Registration of noisy images requires a robust distance measure.

$$\mathrm{NCC}[f,g] = \frac{1}{|\Omega|} \int_\Omega \frac{\left(f-\overline{f}\right)}{\sigma_f} \frac{(g-\overline{g})}{\sigma_g} \,\mathrm{d}x,$$

where

- $\overline{f} = \frac{1}{|\Omega|} \int_\Omega f \,\mathrm{d}x$ (mean value)
- $\sigma_f = \sqrt{\frac{1}{|\Omega|} \int_\Omega \left(f-\overline{f}\right)^2 \mathrm{d}x}$. (standard deviation)

Interpretation

$$\mathrm{NCC}[f,g] = (\tilde{f},\tilde{g})_{L^2} \text{ where } \tilde{f} := \frac{f-\overline{f}}{\left\|f-\overline{f}\right\|} \text{ and } \tilde{g} := \frac{g-\overline{g}}{\|g-\overline{g}\|}$$

Normalized cross-correlation

Registration of noisy images requires a robust distance measure.

$$\mathsf{NCC}[f,g] = \frac{1}{|\Omega|}\int_\Omega \frac{\left(f-\overline{f}\right)}{\sigma_f}\frac{\left(g-\overline{g}\right)}{\sigma_g}\,\mathrm{d}x,$$

where

- $\overline{f} = \frac{1}{|\Omega|}\int_\Omega f\,\mathrm{d}x$ (mean value)
- $\sigma_f = \sqrt{\frac{1}{|\Omega|}\int_\Omega \left(f-\overline{f}\right)^2\mathrm{d}x}$. (standard deviation)

Interpretation

$$\mathsf{NCC}[f,g] = (\tilde{f},\tilde{g})_{L^2} \text{ where } \tilde{f} := \frac{f-\overline{f}}{\left\|f-\overline{f}\right\|} \text{ and } \tilde{g} := \frac{g-\overline{g}}{\|g-\overline{g}\|}$$

Thus, $-1 \leq \mathsf{NCC}[f,g] \leq 1$ and

$$\begin{aligned}\mathsf{NCC}[f,g] = 1 &\Leftrightarrow \tilde{f} = \tilde{g} \text{ a. e.}\\ &\Leftrightarrow f = ag + b \text{ with } a > 0 \text{ and } b \in \mathbb{R}.\end{aligned}$$

Normalized cross-correlation

Registration of noisy images requires a robust distance measure.

$$\mathsf{NCC}[f,g] = \frac{1}{|\Omega|}\int_\Omega \frac{\left(f-\overline{f}\right)}{\sigma_f}\frac{\left(g-\overline{g}\right)}{\sigma_g}\,\mathrm{d}x,$$

where

- $\overline{f} = \frac{1}{|\Omega|}\int_\Omega f\,\mathrm{d}x$ (mean value)
- $\sigma_f = \sqrt{\frac{1}{|\Omega|}\int_\Omega \left(f-\overline{f}\right)^2\mathrm{d}x}.$ (standard deviation)

Interpretation

$$\mathsf{NCC}[f,g] = (\tilde{f},\tilde{g})_{L^2} \text{ where } \tilde{f} := \frac{f-\overline{f}}{\left\|f-\overline{f}\right\|} \text{ and } \tilde{g} := \frac{g-\overline{g}}{\|g-\overline{g}\|}$$

Thus, $-1 \le \mathsf{NCC}[f,g] \le 1$ and

$$\mathsf{NCC}[f,g] = 1 \Leftrightarrow \tilde{f} = \tilde{g} \text{ a. e.}$$

$$\Leftrightarrow f = ag + b \text{ with } a > 0 \text{ and } b \in \mathbb{R}.$$

Data functional $\quad E_{\mathsf{NCC}}[\phi] = -\mathsf{NCC}[g_{\mathsf{R}}, g_{\mathsf{T}} \circ \phi]$

Image reconstruction from an image series

Energy for joint registration and reconstruction

$$E[f,\phi_1,\dots,\phi_n] = \sum_{i=1}^{n}\left(-\mathsf{NCC}[f, f_i\circ\phi_i] + \frac{\lambda}{2}\int_\Omega \|D\phi_i - \mathbb{1}\|^2 \,\mathrm{d}x\right)$$

[B., Binev, Blom, Dahmen, Sharpley, Vogt Ultramicroscopy '14]

Energy for joint registration and reconstruction

$$E[f, \phi_1, \dots, \phi_n] = \sum_{i=1}^{n} \left(-\mathsf{NCC}[f, f_i \circ \phi_i] + \frac{\lambda}{2} \int_\Omega \|D\phi_i - \mathbb{1}\|^2 \, \mathrm{d}x \right)$$

[B., Binev, Blom, Dahmen, Sharpley, Vogt Ultramicroscopy '14]

Theorem Let Ω be a bounded domain and $f_1, \dots, f_n \in C(\mathbb{R}^d)$ bounded. Then, a minimizer of

$$E : L^2(\Omega) \times M^n \to \mathbb{R}$$

exists, where

$$M := \left\{ \phi \in H^{1,2}(\Omega, \mathbb{R}^d) \,:\, \big\| \|\phi - \mathrm{id}\|_2 \big\|_{L^\infty} \leq \mathrm{diam}(\Omega) \right\}.$$

Energy for joint registration and reconstruction

$$E[f,\phi_1,\dots,\phi_n] = \sum_{i=1}^{n} \left(-\mathsf{NCC}[f, f_i \circ \phi_i] + \frac{\lambda}{2} \int_\Omega \|D\phi_i - \mathbb{1}\|^2 \,\mathrm{d}x \right)$$

[B., Binev, Blom, Dahmen, Sharpley, Vogt Ultramicroscopy '14]

Theorem Let Ω be a bounded domain and $f_1,\dots,f_n \in C(\mathbb{R}^d)$ bounded. Then, a minimizer of

$$E : L^2(\Omega) \times M^n \to \mathbb{R}$$

exists, where

$$M := \left\{ \phi \in H^{1,2}(\Omega,\mathbb{R}^d) \;:\; \big\| \|\phi - \mathrm{id}\|_2 \big\|_{L^\infty} \leq \mathrm{diam}(\Omega) \right\}.$$

Numerical minimization strategy

- initialize f with f_1, ϕ_1 with id and ϕ_i, $i > 2$, with $\phi_{i,1}$ (see next slide)
- minimize alternatingly with respect to $f, \phi_1, \dots, \phi_n$

Energy for joint registration and reconstruction

$$E[f,\phi_1,\dots,\phi_n]=\sum_{i=1}^{n}\left(-\mathsf{NCC}[f,f_i\circ\phi_i]+\frac{\lambda}{2}\int_\Omega \|D\phi_i-\mathbb{1}\|^2\,\mathrm{d}x\right)$$

[B., Binev, Blom, Dahmen, Sharpley, Vogt Ultramicroscopy '14]

Theorem Let Ω be a bounded domain and $f_1,\dots,f_n\in C(\mathbb{R}^d)$ bounded. Then, a minimizer of

$$E: L^2(\Omega)\times M^n\to\mathbb{R}$$

exists, where

$$M:=\left\{\phi\in H^{1,2}(\Omega,\mathbb{R}^d)\;:\;\|\|\phi-\mathrm{id}\|_2\|_{L^\infty}\leq \mathrm{diam}(\Omega)\right\}.$$

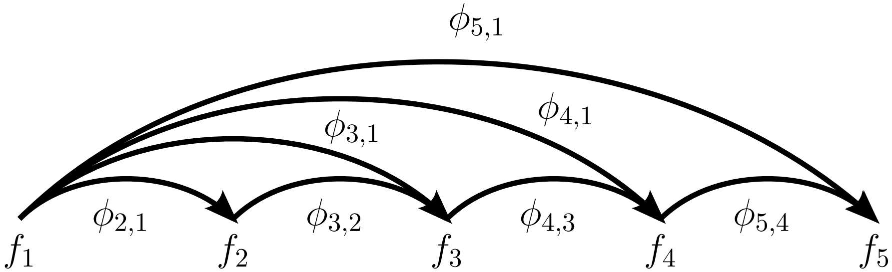
Numerical minimization strategy

- initialize f with f_1, ϕ_1 with id and ϕ_i, $i>2$, with $\phi_{i,1}$ (see next slide)
- minimize alternatingly with respect to $f,\phi_1,\dots,\phi_n$

Note This strategy is biased to f_1. [B., Liebscher Ultramic. '19]

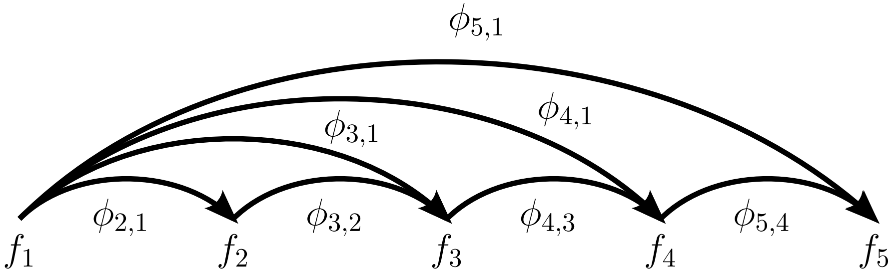
Registering long image series

Strategy



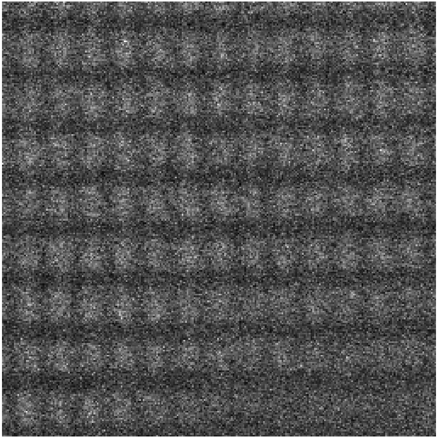
Registering long image series

Strategy

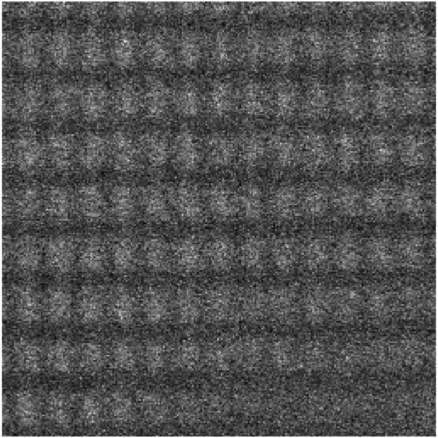


- compute $\phi_{i+1,i}$ for $i = 1, \ldots, n-1$ (initial guess id)
- compute $\phi_{3,1}$ (initial guess $\phi_{3,2} \circ \phi_{2,1}$)
- compute $\phi_{4,1}$ (initial guess $\phi_{4,3} \circ \phi_{3,1}$)
- $\ldots$
- compute $\phi_{n,1}$ (initial guess $\phi_{n,n-1} \circ \phi_{n-1,1}$)

Non-rigid registration – Aligning a series



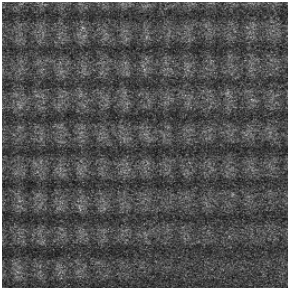
initial data



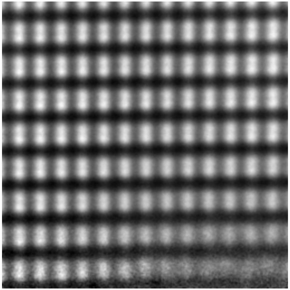
registered data

Silicon (112), magnification 82Mx, dwell time 3.2µs per pixel

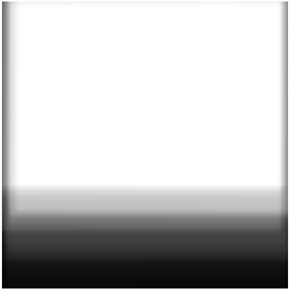
Non-rigid registration – Reconstruction from 512 frames



f_1



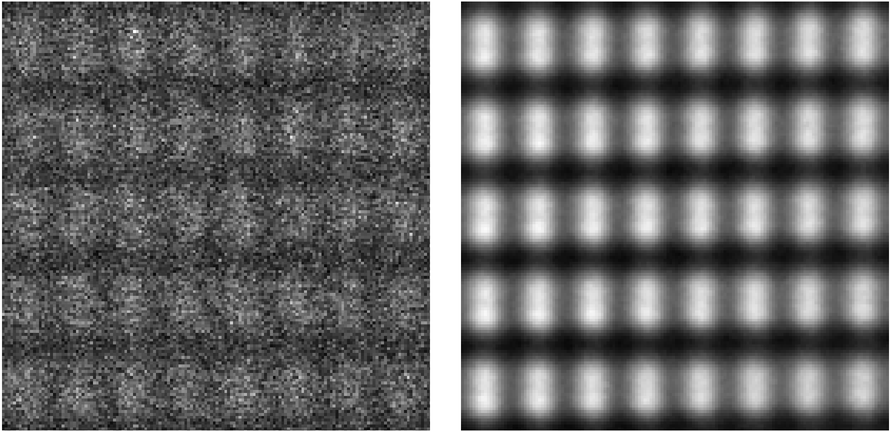
reconstruction



#samples

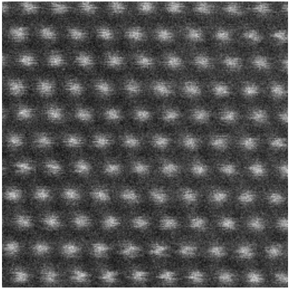
Silicon (112), magnification 82Mx, dwell time 3.2µs

Non-rigid registration – Reconstruction from 512 frames

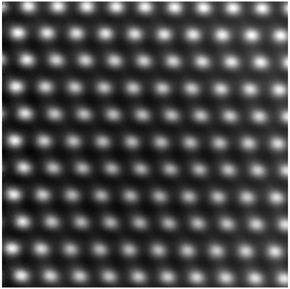


Silicon (112), magnification 82Mx, dwell time 3.2µs

Non-rigid registration – Reconstruction from 512 frames



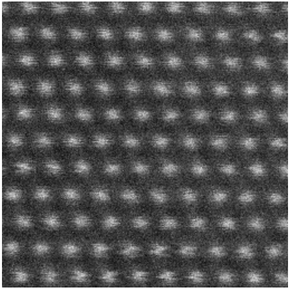
f_1



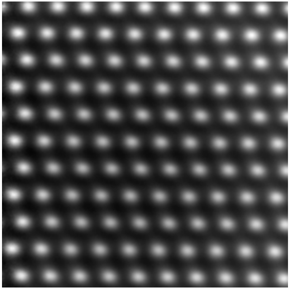
reconstruction

#samples

GaN [11-20], magnification 29Mx, dwell time 12μs



f_1



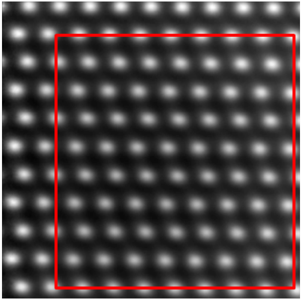
reconstruction

#samples

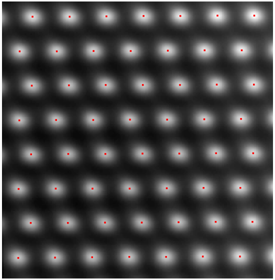
GaN [11-20], magnification 29Mx, dwell time 12μs

How to evaluate the quality of the reconstruction? Ground truth?

Results – Precision analysis on GaN [11-20]

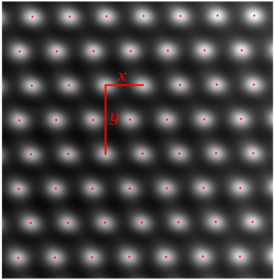


Results – Precision analysis on GaN [11-20]



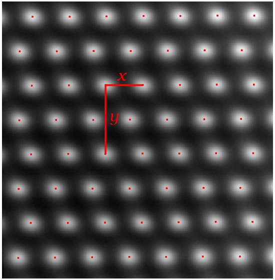
- detect atom centers

Results – Precision analysis on GaN [11-20]



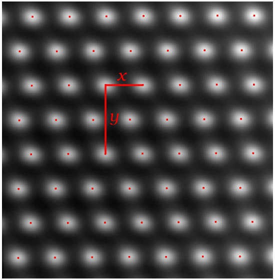
- detect atom centers
- compute marked x- and y-distances for all available atom pairs

Results – Precision analysis on GaN [11-20]



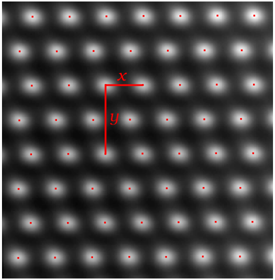
- detect atom centers
- compute marked x- and y-distances for all available atom pairs
- compute standard deviation of these distances (separately for x and y)

Results – Precision analysis on GaN [11-20]



- detect atom centers
- compute marked x- and y-distances for all available atom pairs
- compute standard deviation of these distances (separately for x and y)
- in the context of electron microscopy, this standard deviation is called *precision*

Results – Precision analysis on GaN [11-20]



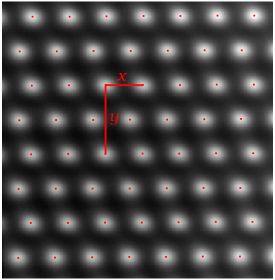
- detect atom centers
- compute marked x- and y-distances for all available atom pairs
- compute standard deviation of these distances (separately for x and y)
- in the context of electron microscopy, this standard deviation is called *precision*

The attained precision for GaN [11-20] (pixel size 21pm) is:

- x-precision $= 0.74$pm
- y-precision $= 0.85$pm

[Yankovich, B., Dahmen, Binev, Sanchez, Bradley, Li, Szlufarska, Voyles Nature Comm. '14]

Results – Precision analysis on GaN [11-20]



- detect atom centers
- compute marked x- and y-distances for all available atom pairs
- compute standard deviation of these distances (separately for x and y)
- in the context of electron microscopy, this standard deviation is called *precision*

The attained precision for GaN [11-20] (pixel size 21pm) is:

- x-precision $= 0.74$pm
- y-precision $= 0.85$pm

[Yankovich, B., Dahmen, Binev, Sanchez, Bradley, Li, Szlufarska, Voyles Nature Comm. '14]

Before best reported precision from STEM series was ~ 5pm.

[Kimoto et al. Ultramicr. '10]

Precision achievable in a single STEM image typically ~ 15pm.

[Schmid et al. Micron '12]

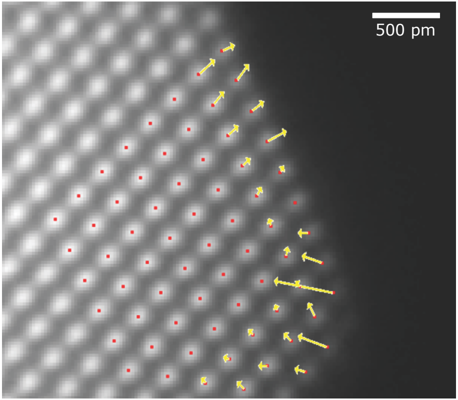
Non-rigid registration - Analysis of a Pt nanocatalyst

High precision allows to analyze defects in the crystal lattice:

[Yankovich, B., Dahmen, Binev, Sanchez, Bradley, Li, Szlufarska, Voyles Nature Comm. '14]

Non-rigid registration - Analysis of a Pt nanocatalyst

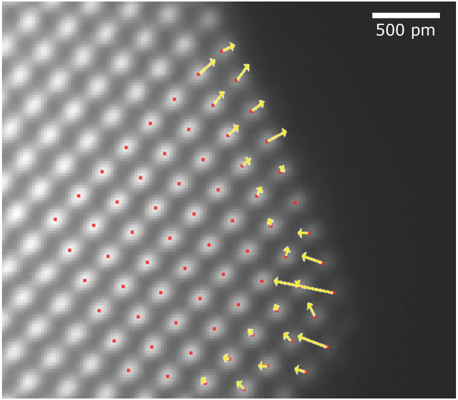
High precision allows to analyze defects in the crystal lattice:



[Yankovich, B., Dahmen, Binev, Sanchez, Bradley, Li, Szlufarska, Voyles Nature Comm. '14]

Non-rigid registration - Analysis of a Pt nanocatalyst

High precision allows to analyze defects in the crystal lattice:

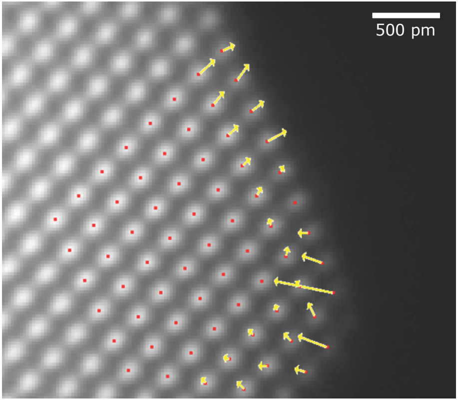


Positions at the nanoparticle corner differ from the regular lattice.

[Yankovich, B., Dahmen, Binev, Sanchez, Bradley, Li, Szlufarska, Voyles Nature Comm. '14]

Non-rigid registration - Analysis of a Pt nanocatalyst

High precision allows to analyze defects in the crystal lattice:



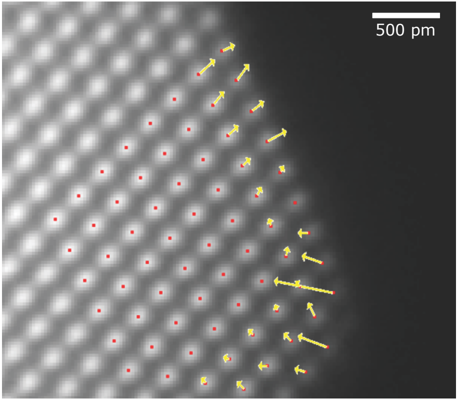
Positions at the nanoparticle corner differ from the regular lattice.

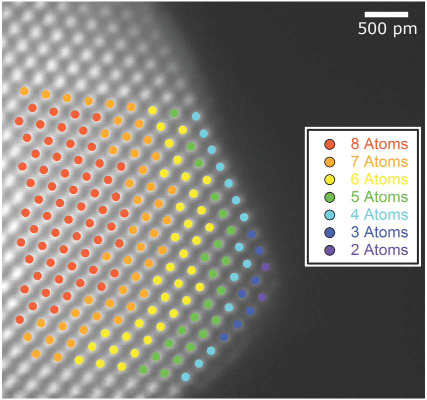
High signal-to-noise ratio allows to estimate the 3D particle structure.

[Yankovich, B., Dahmen, Binev, Sanchez, Bradley, Li, Szlufarska, Voyles Nature Comm. '14]

Non-rigid registration - Analysis of a Pt nanocatalyst

High precision allows to analyze defects in the crystal lattice:



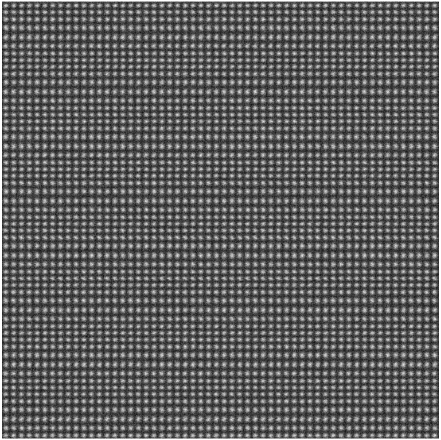


Positions at the nanoparticle corner differ from the regular lattice.

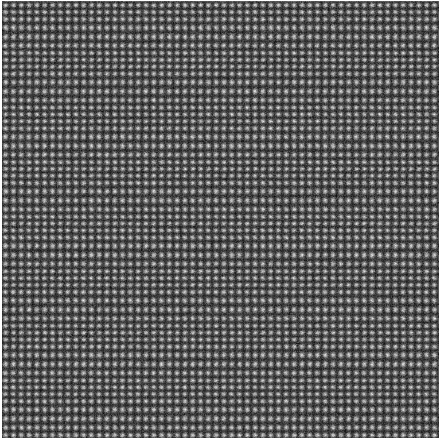
High signal-to-noise ratio allows to estimate the 3D particle structure.

[Yankovich, B., Dahmen, Binev, Sanchez, Bradley, Li, Szlufarska, Voyles Nature Comm. '14]

Limits of alternating minimization - Initialization bias



initial data



registered data

Artificial series with fast and slow scan noise

joint work with C. Liebscher (MPIE Düsseldorf)

Bias correction

Observation: If ψ is a translation, we have

$$E[f, \phi_1, \ldots, \phi_n] = E[f \circ \psi, \phi_1 \circ \psi, \ldots, \phi_n \circ \psi].$$

Bias correction

Observation: If ψ is a translation, we have

$$E[f, \phi_1, \dots, \phi_n] = E[f \circ \psi, \phi_1 \circ \psi, \dots, \phi_n \circ \psi].$$

Idea: Reduce the deformations, i. e. find ψ that minimizes

$$\sum_{i=1}^{n} \int_\Omega \|\phi_i(\psi(x)) - x\|^2 \, \mathrm{d}x \, .$$

Bias correction

Observation: If ψ is a translation, we have

$$E[f, \phi_1, \ldots, \phi_n] = E[f \circ \psi, \phi_1 \circ \psi, \ldots, \phi_n \circ \psi].$$

Idea: Reduce the deformations, i. e. find ψ that minimizes

$$\sum_{i=1}^{n} \int_{\Omega} \|\phi_i(\psi(x)) - x\|^2 \, \mathrm{d}x \, .$$

Note: $f \approx f_i \circ \phi_i \Rightarrow f \circ \psi \approx f_i \circ \phi_i \circ \psi$

Bias correction

Observation: If ψ is a translation, we have

$$E[f, \phi_1, \ldots, \phi_n] = E[f \circ \psi, \phi_1 \circ \psi, \ldots, \phi_n \circ \psi].$$

Idea: Reduce the deformations, i. e. find ψ that minimizes

$$\sum_{i=1}^{n} \int_\Omega \|\phi_i(\psi(x)) - x\|^2 \, \mathrm{d}x \, .$$

Note: $f \approx f_i \circ \phi_i \Rightarrow f \circ \psi \approx f_i \circ \phi_i \circ \psi$

Strategy:

After each alternate minimization step

- compute ψ (with a regularized gradient descent)
- replace ϕ_i by $\phi_i \circ \psi$
- replace f by $f \circ \psi$

Bias correction

Observation: If ψ is a translation, we have

$$E[f, \phi_1, \dots, \phi_n] = E[f \circ \psi, \phi_1 \circ \psi, \dots, \phi_n \circ \psi].$$

Idea: Reduce the deformations, i. e. find ψ that minimizes

$$\sum_{i=1}^{n} \int_{\Omega} \|\phi_i(\psi(x)) - x\|^2 \,\mathrm{d}x \,.$$

Note: $f \approx f_i \circ \phi_i \Rightarrow f \circ \psi \approx f_i \circ \phi_i \circ \psi$

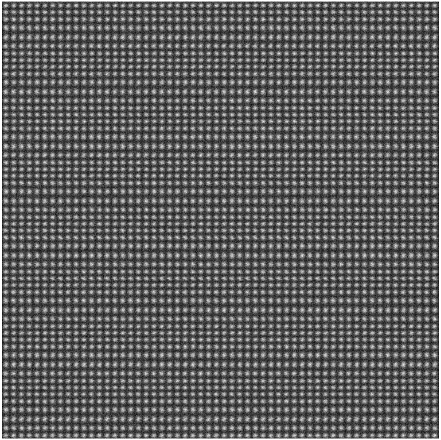
Strategy:

After each alternate minimization step

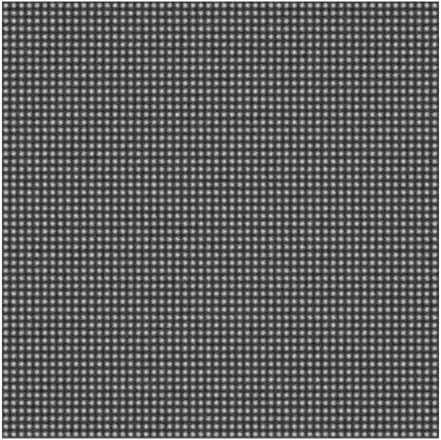
- compute ψ (with a regularized gradient descent)
- replace ϕ_i by $\phi_i \circ \psi$
- replace f by $f \circ \psi$

Note: The computation of ψ introduces a direct coupling of the ϕ_i.

Bias correction - Artificial results



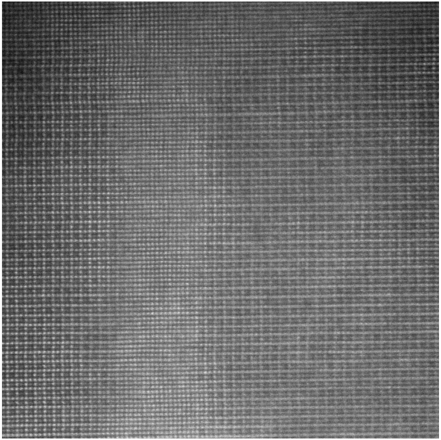
initial data



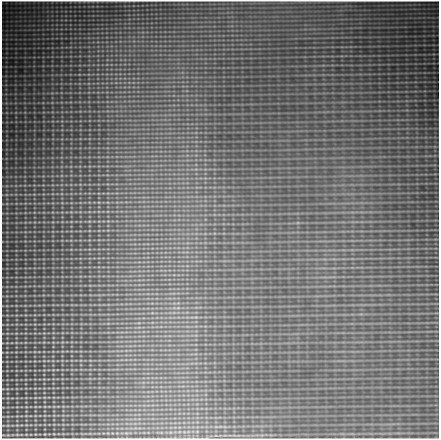
registered data with reduction

[B., Liebscher Ultramic. '19]

Bias correction - Reconstruction from 19 STEM frames



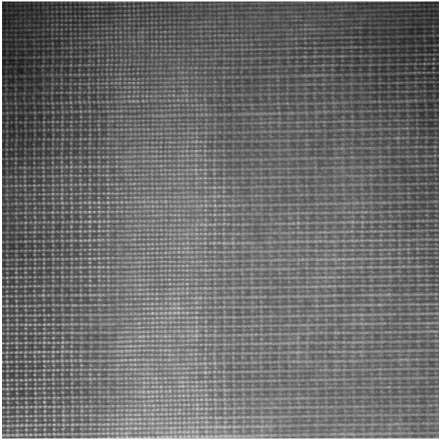
f_1



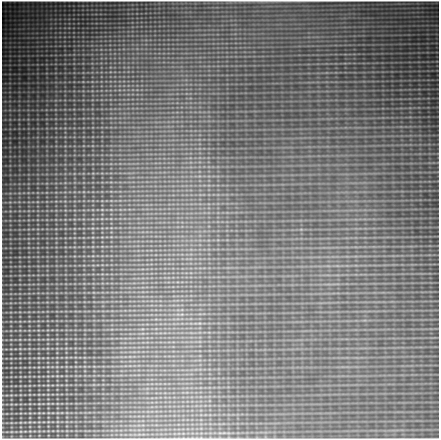
reconstruction without reduction

[B., Liebscher Ultramic. '19]

Bias correction - Reconstruction from 19 STEM frames



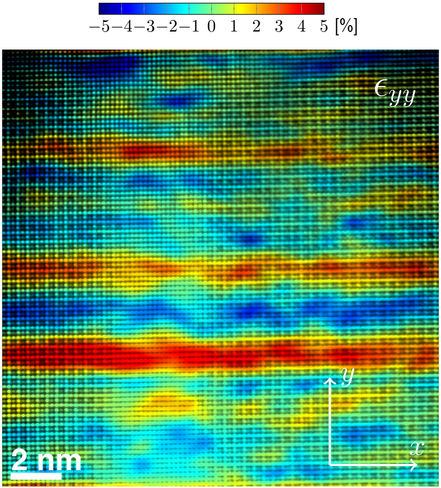
f_1



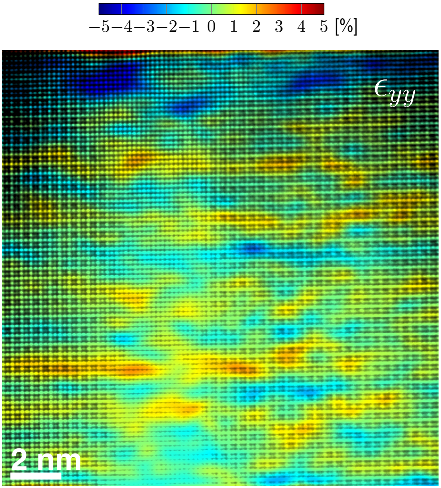
reconstruction with reduction

[B., Liebscher Ultramic. '19]

Bias correction - Effect on strain estimation



no bias-correction



with bias-correction

[B., Liebscher Ultramic. '19]

Source code

The C++ source code of our registration framework is available at

`https://github.com/berkels/match-series`

Source code

The C++ source code of our registration framework is available at

`https://github.com/berkels/match-series`

x86-64 binaries for Linux, Mac and Windows are available via Anaconda:

```
conda install -c conda-forge match-series
```

Thanks to Jan Janßen, MPIE, for the conda package!

Source code

The C++ source code of our registration framework is available at

`https://github.com/berkels/match-series`

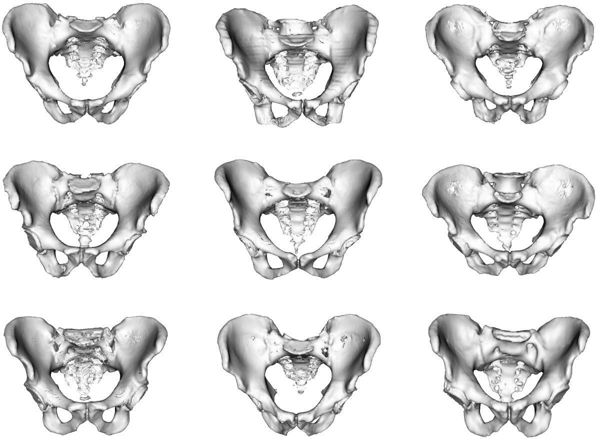
x86-64 binaries for Linux, Mac and Windows are available via Anaconda:

```
conda install -c conda-forge match-series
```

Thanks to Jan Janßen, MPIE, for the conda package!

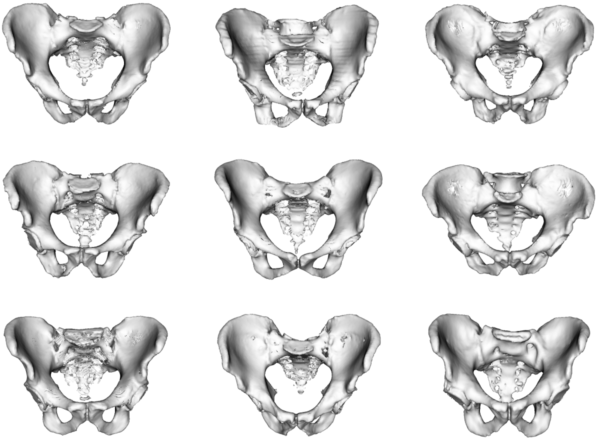
Usage via a command line tool configured via a parameter file.

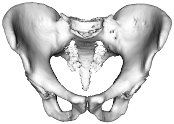
Shape averaging for jawbone reconstruction



joint work with A. Modabber, F. Peters (University Hospital Aachen)

Shape averaging for jawbone reconstruction





average pelvis

joint work with A. Modabber, F. Peters (University Hospital Aachen)

Existence of minimizers and some consequences

Existence of minimizers

J needs to be bounded from below on the admissible set, i.e.

$$\underline{J} := \inf_{x \in X} J[x] > -\infty.$$

(not sufficient, e.g. $J[x] = e^x$, this J is even strictly convex and analytic)
Then, there is a *minimizing sequence*, i.e. $(x_n)_n \in X^{\mathbb{N}}$ with $J[x_n] \to \underline{J}$ for $n \to \infty$.

Existence of minimizers

J needs to be bounded from below on the admissible set, i.e.

$$\underline{J} := \inf_{x \in X} J[x] > -\infty.$$

(not sufficient, e.g. $J[x] = e^x$, this J is even strictly convex and analytic)
Then, there is a *minimizing sequence*, i.e. $(x_n)_n \in X^{\mathbb{N}}$ with $J[x_n] \to \underline{J}$ for $n \to \infty$.

Direct method in the calculus of variations:

1. Selection of a minimizing sequence $(x_n)_n \in X^{\mathbb{N}}$
2. Getting a convergent subsequence $(x_{n_k})_k \in X^{\mathbb{N}}$ (with $x^* \in X$)
3. Proving *lower semi-continuity* of J, i.e.

$$J[y] \leq \liminf_{n \to \infty} J[y_n] \text{ for all } (y_n)_n \in X^{\mathbb{N}} \text{ with } y_n \to y \in X.$$

   This means that function values do not "jump down".

Existence of minimizers

J needs to be bounded from below on the admissible set, i.e.

$$\underline{J} := \inf_{x \in X} J[x] > -\infty.$$

(not sufficient, e.g. $J[x] = e^x$, this J is even strictly convex and analytic)
Then, there is a *minimizing sequence*, i.e. $(x_n)_n \in X^{\mathbb{N}}$ with $J[x_n] \to \underline{J}$ for $n \to \infty$.

Direct method in the calculus of variations:

1. Selection of a minimizing sequence $(x_n)_n \in X^{\mathbb{N}}$
2. Getting a convergent subsequence $(x_{n_k})_k \in X^{\mathbb{N}}$ (with $x^* \in X$)
3. Proving *lower semi-continuity* of J, i.e.

$$J[y] \leq \liminf_{n \to \infty} J[y_n] \text{ for all } (y_n)_n \in X^{\mathbb{N}} \text{ with } y_n \to y \in X.$$

   This means that function values do not "jump down".

Then, x^* is a minimizer, i.e. $J[x^*] = \underline{J}$, since

$$\underline{J} = \lim_{n \to \infty} J[x_n] = \lim_{k \to \infty} J[x_{n_k}] = \liminf_{k \to \infty} J[x_{n_k}] \geq J[x^*] \geq \underline{J}.$$

Existence of minimizers (cont.)

Applying the direct method in case $\dim(X) < \infty$ is rather simple.

Existence of minimizers (cont.)

Applying the direct method in case $\dim(X) < \infty$ is rather simple.

This is mainly due to the following: If $\dim(X) < \infty$,

- all norms are equivalent and
- norm-bounded sequences have a convergent subsequence.

Existence of minimizers (cont.)

Applying the direct method in case $\dim(X) < \infty$ is rather simple.

This is mainly due to the following: If $\dim(X) < \infty$,

- all norms are equivalent and
- norm-bounded sequences have a convergent subsequence.

Both is not true if $\dim(X) = \infty$!

Existence of minimizers (cont.)

Applying the direct method in case $\dim(X) < \infty$ is rather simple.

This is mainly due to the following: If $\dim(X) < \infty$,

- all norms are equivalent and
- norm-bounded sequences have a convergent subsequence.

Both is not true if $\dim(X) = \infty$!

Here, a different notion of convergence is needed, weak convergence.

Existence of minimizers (cont.)

Applying the direct method in case $\dim(X) < \infty$ is rather simple.

This is mainly due to the following: If $\dim(X) < \infty$,

- all norms are equivalent and
- norm-bounded sequences have a convergent subsequence.

Both is not true if $\dim(X) = \infty$!

Here, a different notion of convergence is needed, weak convergence.

$\rightarrow$ **Functional analysis**

Existence of minimizers (cont.)

Applying the direct method in case $\dim(X) < \infty$ is rather simple.

This is mainly due to the following: If $\dim(X) < \infty$,

- all norms are equivalent and
- norm-bounded sequences have a convergent subsequence.

Both is not true if $\dim(X) = \infty$!

Here, a different notion of convergence is needed, weak convergence.

$\rightarrow$ **Functional analysis**

Why should we care? After discretization, we always have $\dim(X) < \infty$.

Existence of minimizers (cont.)

Applying the direct method in case $\dim(X) < \infty$ is rather simple.

This is mainly due to the following: If $\dim(X) < \infty$,

- all norms are equivalent and
- norm-bounded sequences have a convergent subsequence.

Both is not true if $\dim(X) = \infty$!

Here, a different notion of convergence is needed, weak convergence.

$\rightarrow$ **Functional analysis**

Why should we care? After discretization, we always have $\dim(X) < \infty$.

The functional space setting gives us information about inherent properties of the solution, e.g. its regularity, that the numerical solution approximates.

A reason to check the existence of minimizers

With the direct method, one can show that minimizers of

$$J[y] = \frac{1}{2} \|y - g\|_{L^2}^2 + \frac{\lambda}{2} \|\nabla y\|_{L^2}^2 \qquad \text{(simple denoising model)}$$

are in $H^{1,2}(\Omega)$.

A reason to check the existence of minimizers

With the direct method, one can show that minimizers of

$$J[y] = \frac{1}{2} \|y - g\|_{L^2}^2 + \frac{\lambda}{2} \|\nabla y\|_{L^2}^2 \qquad \text{(simple denoising model)}$$

are in $H^{1,2}(\Omega)$. Is $H^{1,2}(\Omega)$ a suitable space for images?

A reason to check the existence of minimizers

With the direct method, one can show that minimizers of

$$J[y] = \frac{1}{2} \|y - g\|_{L^2}^2 + \frac{\lambda}{2} \|\nabla y\|_{L^2}^2 \qquad \text{(simple denoising model)}$$

are in $H^{1,2}(\Omega)$. Is $H^{1,2}(\Omega)$ a suitable space for images?

Images can have edges, e.g. jumps in image intensity.

A reason to check the existence of minimizers

With the direct method, one can show that minimizers of

$$J[y] = \frac{1}{2}\|y - g\|_{L^2}^2 + \frac{\lambda}{2}\|\nabla y\|_{L^2}^2 \qquad \text{(simple denoising model)}$$

are in $H^{1,2}(\Omega)$. Is $H^{1,2}(\Omega)$ a suitable space for images?

Images can have edges, e.g. jumps in image intensity.

Let $\Omega \subset \mathbb{R}^d$ be a bounded domain and $D \subset \mathbb{R}^d$ a bounded domain with piecewise smooth boundary with $\overline{D} \subset \Omega$.

A reason to check the existence of minimizers

With the direct method, one can show that minimizers of

$$J[y] = \frac{1}{2}\,\|y - g\|_{L^2}^2 + \frac{\lambda}{2}\,\|\nabla y\|_{L^2}^2 \qquad \text{(simple denoising model)}$$

are in $H^{1,2}(\Omega)$. Is $H^{1,2}(\Omega)$ a suitable space for images?

Images can have edges, e.g. jumps in image intensity.

Let $\Omega \subset \mathbb{R}^d$ be a bounded domain and $D \subset \mathbb{R}^d$ a bounded domain with piecewise smooth boundary with $\overline{D} \subset \Omega$.
Then, the *characteristic function* χ_D of D, given by

$$\chi_D(x) := \begin{cases} 1 & x \in D \\ 0 & \text{else} \end{cases},$$

is not in $H^{1,p}(\Omega)$ for $1 \leq p \leq \infty$.

A reason to check the existence of minimizers

With the direct method, one can show that minimizers of

$$J[y] = \frac{1}{2}\,\|y - g\|_{L^2}^2 + \frac{\lambda}{2}\,\|\nabla y\|_{L^2}^2 \qquad \text{(simple denoising model)}$$

are in $H^{1,2}(\Omega)$. Is $H^{1,2}(\Omega)$ a suitable space for images?

Images can have edges, e.g. jumps in image intensity.

Let $\Omega \subset \mathbb{R}^d$ be a bounded domain and $D \subset \mathbb{R}^d$ a bounded domain with piecewise smooth boundary with $\overline{D} \subset \Omega$.
Then, the *characteristic function* χ_D of D, given by

$$\chi_D(x) := \begin{cases} 1 & x \in D \\ 0 & \text{else} \end{cases},$$

is not in $H^{1,p}(\Omega)$ for $1 \leq p \leq \infty$.

This means that the simple denoising model cannot preserve edges!

A suitable space for denoised images

Let $y \in C^1[0,1]$ be increasing. Then,

$$|y|_{H^{1,1}} = \int_0^1 \left|y'(t)\right| \mathrm{d}t = \int_0^1 y'(t)\,\mathrm{d}t = y(1) - y(0).$$

Thus, $|y|_{H^{1,1}}$ is independent of $y'(t)$, just $y(1) - y(0)$ matters.

A suitable space for denoised images

Let $y \in C^1[0,1]$ be increasing. Then,

$$|y|_{H^{1,1}} = \int_0^1 \left|y'(t)\right| \mathrm{d}t = \int_0^1 y'(t)\,\mathrm{d}t = y(1) - y(0).$$

Thus, $|y|_{H^{1,1}}$ is independent of $y'(t)$, just $y(1) - y(0)$ matters.

In particular, a function with jump like

$$(0,1) \to \mathbb{R}, t \mapsto \begin{cases} 0 & t < \frac{1}{2} \\ 1 & t \geq \frac{1}{2} \end{cases}$$

can be approximated with a sequence bounded in the $H^{1,1}$-norm.

A suitable space for denoised images

Let $y \in C^1[0,1]$ be increasing. Then,

$$|y|_{H^{1,1}} = \int_0^1 \left|y'(t)\right| \mathrm{d}t = \int_0^1 y'(t)\,\mathrm{d}t = y(1) - y(0).$$

Thus, $|y|_{H^{1,1}}$ is independent of $y'(t)$, just $y(1) - y(0)$ matters.

In particular, a function with jump like

$$(0,1) \to \mathbb{R}, t \mapsto \begin{cases} 0 & t < \frac{1}{2} \\ 1 & t \geq \frac{1}{2} \end{cases}$$

can be approximated with a sequence bounded in the $H^{1,1}$-norm.

Still, we need to extend the $H^{1,1}$-norm to such functions, we need a more general concept than weak derivatives.

A suitable space for denoised images (cont.)

For $x \in \mathbb{R}^d$ with $x \neq 0$, we have

$$\|x\|_2 = x \cdot \frac{x}{\|x\|_2} \leq \sup_{\|p\|_2 \leq 1} -x \cdot p \leq \sup_{\|p\|_2 \leq 1} \|x\|_2 \|p\|_2 = \|x\|_2$$

$$\Rightarrow \|x\|_2 = \sup_{\|p\|_2 \leq 1} -x \cdot p.$$

A suitable space for denoised images (cont.)

For $x \in \mathbb{R}^d$ with $x \neq 0$, we have

$$\|x\|_2 = x \cdot \frac{x}{\|x\|_2} \leq \sup_{\|p\|_2 \leq 1} -x \cdot p \leq \sup_{\|p\|_2 \leq 1} \|x\|_2 \|p\|_2 = \|x\|_2$$

$$\Rightarrow \|x\|_2 = \sup_{\|p\|_2 \leq 1} -x \cdot p.$$

Thus, for

$$K = \left\{ p \in C_c^\infty(\Omega, \mathbb{R}^d) \ : \ \|p(x)\|_2 \leq 1 \text{ for all } x \in \Omega \right\}$$

and $y \in H^{1,1}(\Omega)$, we have

$$\int_\Omega \|\nabla y\|_2 \,\mathrm{d}x = \sup_{p \in K} \int_\Omega -\nabla y \cdot p \,\mathrm{d}x = \sup_{p \in K} \left(- \int_\Omega \sum_{i=1}^d \partial_i y p_i \,\mathrm{d}x \right)$$

$$= \sup_{p \in K} \int_\Omega \sum_{i=1}^d y \partial_i p_i \,\mathrm{d}x = \sup_{p \in K} \int_\Omega y \,\mathrm{div} p \,\mathrm{d}x \,.$$

A suitable space for denoised images (cont.)

For $x \in \mathbb{R}^d$ with $x \neq 0$, we have

$$\|x\|_2 = x \cdot \frac{x}{\|x\|_2} \leq \sup_{\|p\|_2 \leq 1} -x \cdot p \leq \sup_{\|p\|_2 \leq 1} \|x\|_2 \|p\|_2 = \|x\|_2$$

$$\Rightarrow \|x\|_2 = \sup_{\|p\|_2 \leq 1} -x \cdot p.$$

Thus, for

$$K = \left\{ p \in C_c^\infty(\Omega, \mathbb{R}^d) \ : \ \|p(x)\|_2 \leq 1 \text{ for all } x \in \Omega \right\}$$

and $y \in H^{1,1}(\Omega)$, we have

$$\int_\Omega \|\nabla y\|_2 \,\mathrm{d}x = \sup_{p \in K} \int_\Omega -\nabla y \cdot p \,\mathrm{d}x = \sup_{p \in K} \left(- \int_\Omega \sum_{i=1}^d \partial_i y p_i \,\mathrm{d}x \right)$$

$$= \sup_{p \in K} \int_\Omega \sum_{i=1}^d y \partial_i p_i \,\mathrm{d}x = \sup_{p \in K} \int_\Omega y \,\mathrm{div} p \,\mathrm{d}x \,.$$

The latter only needs $y \in L^1(\Omega)$ to be defined.

Functions of bounded variation

For $y \in L^1(\Omega)$, the *total variation* is defined as

$$|y|_{BV(\Omega)} = \sup_{p \in C_c^\infty(\Omega,\mathbb{R}^d) \wedge \left\| \|p\|_2 \right\|_{L^\infty} \leq 1} \int_\Omega y \, \mathrm{div} p \, \mathrm{d}x \, .$$

Functions of bounded variation

For $y \in L^1(\Omega)$, the *total variation* is defined as

$$|y|_{BV(\Omega)} = \sup_{p \in C_c^\infty(\Omega, \mathbb{R}^d) \wedge \left\| \|p\|_2 \right\|_{L^\infty} \leq 1} \int_\Omega y \, \mathrm{div} p \, \mathrm{d}x \, .$$

The space of functions of bounded variation is

$$BV(\Omega) := \left\{ y \in L^1(\Omega) \, : \, |y|_{BV(\Omega)} < \infty \right\} .$$

Functions of bounded variation

For $y \in L^1(\Omega)$, the *total variation* is defined as

$$|y|_{BV(\Omega)} = \sup_{p \in C_c^\infty(\Omega,\mathbb{R}^d) \wedge \big\| \|p\|_2 \big\|_{L^\infty} \leq 1} \int_\Omega y \, \mathrm{div} p \, \mathrm{d}x \, .$$

The space of functions of bounded variation is

$$BV(\Omega) := \Big\{ y \in L^1(\Omega) \, : \, |y|_{BV(\Omega)} < \infty \Big\} \, .$$

The BV-norm of $y \in BV(\Omega)$ is defined as

$$\|y\|_{BV(\Omega)} := \|y\|_{L^1(\Omega)} + |y|_{BV(\Omega)} \, .$$

Functions of bounded variation

For $y \in L^1(\Omega)$, the *total variation* is defined as

$$|y|_{BV(\Omega)} = \sup_{p \in C_c^\infty(\Omega,\mathbb{R}^d) \wedge \left\| \|p\|_2 \right\|_{L^\infty} \leq 1} \int_\Omega y \operatorname{div} p \,\mathrm{d}x \,.$$

The space of functions of bounded variation is

$$BV(\Omega) := \left\{ y \in L^1(\Omega) \;:\; |y|_{BV(\Omega)} < \infty \right\}.$$

The BV-norm of $y \in BV(\Omega)$ is defined as

$$\|y\|_{BV(\Omega)} := \|y\|_{L^1(\Omega)} + |y|_{BV(\Omega)} \,.$$

Let $y \in H^{1,1}(\Omega)$. We have $\int_\Omega \|\nabla y\|_2 \,\mathrm{d}x = |y|_{BV}$. From

$$\|x\|_2 \leq \|x\|_1 \leq \sqrt{d}\,\|x\|_2 \text{ and } |y|_{H^{1,1}} = \int_\Omega \|\nabla y\|_1 \,\mathrm{d}x,$$

it follows that

$$|y|_{BV} \leq |y|_{H^{1,1}} \leq \sqrt{d}\,|y|_{BV} \Rightarrow \|y\|_{BV} \leq \|y\|_{H^{1,1}} \leq \sqrt{d}\,\|y\|_{BV} \,.$$

Functions of bounded variation

For $y \in L^1(\Omega)$, the *total variation* is defined as

$$|y|_{BV(\Omega)} = \sup_{p \in C_c^\infty(\Omega, \mathbb{R}^d) \wedge \left\| \|p\|_2 \right\|_{L^\infty} \leq 1} \int_\Omega y \operatorname{div} p \, \mathrm{d}x \,.$$

The space of functions of bounded variation is

$$BV(\Omega) := \left\{ y \in L^1(\Omega) \; : \; |y|_{BV(\Omega)} < \infty \right\}.$$

The BV-norm of $y \in BV(\Omega)$ is defined as

$$\|y\|_{BV(\Omega)} := \|y\|_{L^1(\Omega)} + |y|_{BV(\Omega)} \,.$$

Let $y \in H^{1,1}(\Omega)$. We have $\int_\Omega \|\nabla y\|_2 \, \mathrm{d}x = |y|_{BV}$. From

$$\|x\|_2 \leq \|x\|_1 \leq \sqrt{d} \, \|x\|_2 \quad \text{and} \quad |y|_{H^{1,1}} = \int_\Omega \|\nabla y\|_1 \, \mathrm{d}x,$$

it follows that

$$|y|_{BV} \leq |y|_{H^{1,1}} \leq \sqrt{d} \, |y|_{BV} \Rightarrow \|y\|_{BV} \leq \|y\|_{H^{1,1}} \leq \sqrt{d} \, \|y\|_{BV} \,.$$

In particular, we have $H^{1,1}(\Omega) \subset BV(\Omega) \subset L^1(\Omega)$.

Total variation examples

Let $p \in C_c^\infty(\mathbb{R})$ with $\|p\|_{L^\infty} \leq 1$.

- For the *Heaviside Function*

$$H : \mathbb{R} \to \mathbb{R}, t \mapsto \begin{cases} 1 & t > 0 \\ 0 & t \leq 0 \end{cases}$$

  it holds that $|H|_{BV(\mathbb{R})} = 1$.

Total variation examples

Let $p \in C_c^\infty(\mathbb{R})$ with $\|p\|_{L^\infty} \leq 1$.

- For the *Heaviside Function*

$$H : \mathbb{R} \to \mathbb{R}, t \mapsto \begin{cases} 1 & t > 0 \\ 0 & t \leq 0 \end{cases}$$

it holds that $|H|_{BV(\mathbb{R})} = 1$. Then,

$$\int_{\mathbb{R}} H(t) \operatorname{div} p(t) \,\mathrm{d}t = \int_0^\infty p'(t) \,\mathrm{d}t = \lim_{n\to\infty} \int_0^n p'(t) \,\mathrm{d}t$$
$$= \lim_{n\to\infty} (p(n) - p(0)) = -p(0) \leq 1.$$

Since there is an admissible p with $p(0) = -1$, we get $|H|_{BV(\mathbb{R})} = 1$.

Total variation examples

Let $p \in C_c^\infty(\mathbb{R})$ with $\|p\|_{L^\infty} \leq 1$.

- For the *Heaviside Function*

$$H : \mathbb{R} \to \mathbb{R}, t \mapsto \begin{cases} 1 & t > 0 \\ 0 & t \leq 0 \end{cases}$$

  it holds that $|H|_{BV(\mathbb{R})} = 1$. Then,

$$\begin{aligned} \int_{\mathbb{R}} H(t)\,\mathrm{div}p(t)\,\mathrm{d}t &= \int_0^\infty p'(t)\,\mathrm{d}t = \lim_{n\to\infty} \int_0^n p'(t)\,\mathrm{d}t \\ &= \lim_{n\to\infty} (p(n) - p(0)) = -p(0) \leq 1. \end{aligned}$$

  Since there is an admissible p with $p(0) = -1$, we get $|H|_{BV(\mathbb{R})} = 1$.
- Let $a, b \in \mathbb{R}$ with $a < b$. Then, $\left|\chi_{[a,b]}\right|_{BV(\mathbb{R})} = 2$.

Total variation examples

Let $p \in C_c^\infty(\mathbb{R})$ with $\|p\|_{L^\infty} \leq 1$.

- For the *Heaviside Function*

$$H : \mathbb{R} \to \mathbb{R}, t \mapsto \begin{cases} 1 & t > 0 \\ 0 & t \leq 0 \end{cases}$$

it holds that $|H|_{BV(\mathbb{R})} = 1$. Then,

$$\begin{aligned} \int_{\mathbb{R}} H(t)\,\mathrm{div}p(t)\,\mathrm{d}t &= \int_0^\infty p'(t)\,\mathrm{d}t = \lim_{n\to\infty} \int_0^n p'(t)\,\mathrm{d}t \\ &= \lim_{n\to\infty} (p(n) - p(0)) = -p(0) \leq 1. \end{aligned}$$

Since there is an admissible p with $p(0) = -1$, we get $|H|_{BV(\mathbb{R})} = 1$.

- Let $a, b \in \mathbb{R}$ with $a < b$. Then, $\left|\chi_{[a,b]}\right|_{BV(\mathbb{R})} = 2$. Then,

$$\int_{\mathbb{R}} \chi_{[a,b]}(t)\,\mathrm{div}p(t)\,\mathrm{d}t = \int_a^b p'(t)\,\mathrm{d}t = p(b) - p(a) \leq 2.$$

Since there is an admissible p with $p(a) = -1$ and $p(b) = 1$, the statement follows.

Total variation examples (cont.)

Consider $\chi_{B_r(0)}$ in $\mathbb{R}^2$. For $p \in C_c^\infty(\mathbb{R}^2, \mathbb{R}^2)$ with $\| \|p\|_2 \|_{L^\infty} \leq 1$, we get

$$\int_{\mathbb{R}^2} \chi_{B_r(0)}(x) \,\mathrm{div} p(x) \,\mathrm{d}x = \int_{B_r(0)} \mathrm{div} p(x) \,\mathrm{d}x = \int_{\partial B_r(0)} p(x) \cdot \nu(x) \,\mathrm{d}A(x)$$

$$\leq \int_{\partial B_r(0)} \|p(x)\|_2 \, \|\nu(x)\|_2 \,\mathrm{d}A(x) \leq \int_{\partial B_r(0)} \mathrm{d}A(x) = 2\pi r.$$

Total variation examples (cont.)

Consider $\chi_{B_r(0)}$ in $\mathbb{R}^2$. For $p \in C_c^\infty(\mathbb{R}^2, \mathbb{R}^2)$ with $\| \|p\|_2 \|_{L^\infty} \leq 1$, we get

$$\begin{aligned}\int_{\mathbb{R}^2} \chi_{B_r(0)}(x)\, \mathrm{div} p(x)\, \mathrm{d}x &= \int_{B_r(0)} \mathrm{div} p(x)\, \mathrm{d}x = \int_{\partial B_r(0)} p(x) \cdot \nu(x)\, \mathrm{d}A(x) \\ &\leq \int_{\partial B_r(0)} \|p(x)\|_2 \, \|\nu(x)\|_2 \, \mathrm{d}A(x) \leq \int_{\partial B_r(0)} \mathrm{d}A(x) = 2\pi r.\end{aligned}$$

Since there is an admissible p with $p = \nu$ on $\partial B_r(0)$, we get

$$\left|\chi_{B_r(0)}\right|_{BV(\mathbb{R}^2)} = 2\pi r.$$

Total variation examples (cont.)

Consider $\chi_{B_r(0)}$ in $\mathbb{R}^2$. For $p \in C_c^\infty(\mathbb{R}^2, \mathbb{R}^2)$ with $\| \|p\|_2 \|_{L^\infty} \leq 1$, we get

$$\int_{\mathbb{R}^2} \chi_{B_r(0)}(x) \operatorname{div} p(x) \, \mathrm{d}x = \int_{B_r(0)} \operatorname{div} p(x) \, \mathrm{d}x = \int_{\partial B_r(0)} p(x) \cdot \nu(x) \, \mathrm{d}A(x)$$
$$\leq \int_{\partial B_r(0)} \|p(x)\|_2 \, \|\nu(x)\|_2 \, \mathrm{d}A(x) \leq \int_{\partial B_r(0)} \mathrm{d}A(x) = 2\pi r.$$

Since there is an admissible p with $p = \nu$ on $\partial B_r(0)$, we get

$$\left|\chi_{B_r(0)}\right|_{BV(\mathbb{R}^2)} = 2\pi r.$$

In general, for a bounded domain $D \subset \mathbb{R}^d$ with piecewise smooth boundary and $\overline{D} \subset \Omega$, $|\chi_D|_{BV(\Omega)}$ is the length of the boundary of D.

Total variation examples (cont.)

Consider $\chi_{B_r(0)}$ in $\mathbb{R}^2$. For $p \in C_c^\infty(\mathbb{R}^2, \mathbb{R}^2)$ with $\|\|p\|_2\|_{L^\infty} \leq 1$, we get

$$\begin{aligned}\int_{\mathbb{R}^2} \chi_{B_r(0)}(x) \, \mathrm{div} p(x) \, \mathrm{d}x &= \int_{B_r(0)} \mathrm{div} p(x) \, \mathrm{d}x = \int_{\partial B_r(0)} p(x) \cdot \nu(x) \, \mathrm{d}A(x) \\ &\leq \int_{\partial B_r(0)} \|p(x)\|_2 \, \|\nu(x)\|_2 \, \mathrm{d}A(x) \leq \int_{\partial B_r(0)} \mathrm{d}A(x) = 2\pi r.\end{aligned}$$

Since there is an admissible p with $p = \nu$ on $\partial B_r(0)$, we get

$$\left|\chi_{B_r(0)}\right|_{BV(\mathbb{R}^2)} = 2\pi r.$$

In general, for a bounded domain $D \subset \mathbb{R}^d$ with piecewise smooth boundary and $\overline{D} \subset \Omega$, $|\chi_D|_{BV(\Omega)}$ is the length of the boundary of D.

In particular, this motivates $\mathrm{Per}(D) := |\chi_D|_{BV(\Omega)}$

Total variation examples (cont.)

Consider $\chi_{B_r(0)}$ in $\mathbb{R}^2$. For $p \in C_c^\infty(\mathbb{R}^2, \mathbb{R}^2)$ with $\| \|p\|_2 \|_{L^\infty} \leq 1$, we get

$$\int_{\mathbb{R}^2} \chi_{B_r(0)}(x) \,\mathrm{div} p(x) \,\mathrm{d}x = \int_{B_r(0)} \mathrm{div} p(x) \,\mathrm{d}x = \int_{\partial B_r(0)} p(x) \cdot \nu(x) \,\mathrm{d}A(x)$$
$$\leq \int_{\partial B_r(0)} \|p(x)\|_2 \,\|\nu(x)\|_2 \,\mathrm{d}A(x) \leq \int_{\partial B_r(0)} \mathrm{d}A(x) = 2\pi r.$$

Since there is an admissible p with $p = \nu$ on $\partial B_r(0)$, we get

$$\left|\chi_{B_r(0)}\right|_{BV(\mathbb{R}^2)} = 2\pi r.$$

In general, for a bounded domain $D \subset \mathbb{R}^d$ with piecewise smooth boundary and $\overline{D} \subset \Omega$, $|\chi_D|_{BV(\Omega)}$ is the length of the boundary of D.

In particular, this motivates $\mathrm{Per}(D) := |\chi_D|_{BV(\Omega)}$

Thus, as regularizer, $|\cdot|_{BV}$ smoothens the boundary of the sublevel sets $\{x \in \Omega \,:\, y(x) < c\}$.

The ROF-functional (Rudin, Osher, Fatemi, 1992)

To denoise an image $g \in L^2(\Omega)$, we are looking for a minimizer of

$$J_{\mathsf{ROF}} : BV(\Omega) \to \mathbb{R}_\infty := \mathbb{R} \cup \{\infty\}, y \mapsto J[y] = \frac{1}{2} \|y - g\|_{L^2}^2 + \lambda |y|_{BV} .$$

The ROF-functional (Rudin, Osher, Fatemi, 1992)

To denoise an image $g \in L^2(\Omega)$, we are looking for a minimizer of

$$J_{\mathsf{ROF}} : BV(\Omega) \to \mathbb{R}_\infty := \mathbb{R} \cup \{\infty\}, y \mapsto J[y] = \frac{1}{2} \|y - g\|_{L^2}^2 + \lambda \, |y|_{BV} \, .$$

Existence of minimizers can be shown with the direct method.

The ROF-functional (Rudin, Osher, Fatemi, 1992)

To denoise an image $g \in L^2(\Omega)$, we are looking for a minimizer of

$$J_{\mathsf{ROF}} : BV(\Omega) \to \mathbb{R}_\infty := \mathbb{R} \cup \{\infty\}, y \mapsto J[y] = \frac{1}{2} \|y - g\|_{L^2}^2 + \lambda |y|_{BV} .$$

Existence of minimizers can be shown with the direct method.

Since,

$$\inf_{u \in BV} J_{\mathsf{ROF}}[u] = \inf_{u \in BV} \sup_{p \in K} \left(\int_\Omega (u - f)^2 + \lambda u \,\mathrm{div} p \,\mathrm{d}x \right),$$

The ROF-functional (Rudin, Osher, Fatemi, 1992)

To denoise an image $g \in L^2(\Omega)$, we are looking for a minimizer of

$$J_{\mathsf{ROF}} : BV(\Omega) \to \mathbb{R}_\infty := \mathbb{R} \cup \{\infty\}, y \mapsto J[y] = \frac{1}{2} \|y - g\|_{L^2}^2 + \lambda \, |y|_{BV} \, .$$

Existence of minimizers can be shown with the direct method.

Since,

$$\inf_{u \in BV} J_{\mathsf{ROF}}[u] = \inf_{u \in BV} \sup_{p \in K} \left(\int_\Omega (u - f)^2 + \lambda u \, \mathrm{div} p \, \mathrm{d}x \right),$$

J_{ROF} can be minimized by solving a saddle point problem.

The ROF-functional (Rudin, Osher, Fatemi, 1992)

To denoise an image $g \in L^2(\Omega)$, we are looking for a minimizer of

$$J_{\mathsf{ROF}} : BV(\Omega) \to \mathbb{R}_\infty := \mathbb{R} \cup \{\infty\}, y \mapsto J[y] = \frac{1}{2} \|y - g\|_{L^2}^2 + \lambda |y|_{BV} .$$

Existence of minimizers can be shown with the direct method.

Since,

$$\inf_{u \in BV} J_{\mathsf{ROF}}[u] = \inf_{u \in BV} \sup_{p \in K} \left(\int_\Omega (u - f)^2 + \lambda u \, \mathrm{div} p \, \mathrm{d}x \right),$$

J_{ROF} can be minimized by solving a saddle point problem.

To derive minimization algorithms, we first *characterize* minimizers.

The ROF-functional (Rudin, Osher, Fatemi, 1992)

To denoise an image $g \in L^2(\Omega)$, we are looking for a minimizer of

$$J_{\mathsf{ROF}} : BV(\Omega) \to \mathbb{R}_\infty := \mathbb{R} \cup \{\infty\}, y \mapsto J[y] = \frac{1}{2} \|y - g\|_{L^2}^2 + \lambda \, |y|_{BV} \, .$$

Existence of minimizers can be shown with the direct method.

Since,

$$\inf_{u \in BV} J_{\mathsf{ROF}}[u] = \inf_{u \in BV} \sup_{p \in K} \left(\int_\Omega (u - f)^2 + \lambda u \, \mathrm{div} p \, \mathrm{d}x \right) ,$$

J_{ROF} can be minimized by solving a saddle point problem.

To derive minimization algorithms, we first *characterize* minimizers.

Note: $|\cdot|_{BV}$ is not differentiable, but convex.

Characterization of minimizers in classical optimization

Necessary condition Let $X = \mathbb{R}^d$, $M \subset X$ open and $J \in C^1(M)$. If $y^* \in M \subset X$ is a local extremum of J, then

$$\nabla J(y^*) = 0.$$

Characterization of minimizers in classical optimization

Necessary condition Let $X = \mathbb{R}^d$, $M \subset X$ open and $J \in C^1(M)$. If $y^* \in M \subset X$ is a local extremum of J, then

$$\nabla J(y^*) = 0.$$

Proposition Let $M \subset \mathbb{R}^d$ be convex and open, and $J \in C^1(M)$. Then,

1. J convex on $M \Leftrightarrow \forall x, y \in M : J(y) \geq J(x) + \nabla J(x) \cdot (y - x)$
2. J strictly convex on M
   $\Leftrightarrow \forall x, y \in M, x \neq y : J(y) > J(x) + \nabla J(x) \cdot (y - x)$

Necessary condition Let $X = \mathbb{R}^d$, $M \subset X$ open and $J \in C^1(M)$. If $y^* \in M \subset X$ is a local extremum of J, then

$$\nabla J(y^*) = 0.$$

Proposition Let $M \subset \mathbb{R}^d$ be convex and open, and $J \in C^1(M)$. Then,

1. J convex on $M \Leftrightarrow \forall x, y \in M : J(y) \geq J(x) + \nabla J(x) \cdot (y - x)$
2. J strictly convex on M
   $\Leftrightarrow \forall x, y \in M, x \neq y : J(y) > J(x) + \nabla J(x) \cdot (y - x)$

Corollary Let M and J be as above and J additionally convex. Moreover, let $y^* \in M$ with $\nabla J[y^*] = 0$. Then,

$$y^* \in \underset{y \in M}{\operatorname{argmin}}\, J[y].$$

Characterization of minimizers in classical optimization

Necessary condition Let $X = \mathbb{R}^d$, $M \subset X$ open and $J \in C^1(M)$. If $y^* \in M \subset X$ is a local extremum of J, then

$$\nabla J(y^*) = 0.$$

Proposition Let $M \subset \mathbb{R}^d$ be convex and open, and $J \in C^1(M)$. Then,

1. J convex on $M \Leftrightarrow \forall x, y \in M : J(y) \geq J(x) + \nabla J(x) \cdot (y - x)$
2. J strictly convex on M
   $\Leftrightarrow \forall x, y \in M, x \neq y : J(y) > J(x) + \nabla J(x) \cdot (y - x)$

Corollary Let M and J be as above and J additionally convex. Moreover, let $y^* \in M$ with $\nabla J[y^*] = 0$. Then,

$$y^* \in \operatorname*{argmin}_{y \in M} J[y].$$

Proof Let $y \in M$ be arbitrary but fixed. Then,

$$J[y] \geq J[y^*] + \nabla J[y^*] \cdot (y - y^*) = J[y^*].$$

□

Sufficient conditions

Let X be a normed vector space and $J : X \to \mathbb{R}_\infty$ convex. $u \in X'$ is called *subgradient* of J, if

$$J[y] + \langle u, x - y \rangle \leq J[x] \text{ for all } x \in X.$$

$\partial J[y]$, the set of all subgradients of J at y, is called *subdifferential* of J.

Sufficient conditions

Let X be a normed vector space and $J : X \to \mathbb{R}_\infty$ convex. $u \in X'$ is called *subgradient* of J, if

$$J[y] + \langle u, x - y \rangle \leq J[x] \text{ for all } x \in X.$$

$\partial J[y]$, the set of all subgradients of J at y, is called *subdifferential* of J.

Proposition Let X be a normed vector space and $J : X \to \mathbb{R}_\infty$ convex. Then,

$$y^* \in \operatorname*{argmin}_{y \in X} J[y] \;\Leftrightarrow\; 0 \in \partial J[y^*].$$

Sufficient conditions

Let X be a normed vector space and $J : X \to \mathbb{R}_\infty$ convex. $u \in X'$ is called *subgradient* of J, if

$$J[y] + \langle u, x - y\rangle \leq J[x] \text{ for all } x \in X.$$

$\partial J[y]$, the set of all subgradients of J at y, is called *subdifferential* of J.

Proposition Let X be a normed vector space and $J : X \to \mathbb{R}_\infty$ convex. Then,

$$y^* \in \operatorname*{argmin}_{y \in X} J[y] \;\Leftrightarrow\; 0 \in \partial J[y^*].$$

Proof y^* is a minimizer, if and only if $J[y^*] \leq J[y]$ for all $y \in X$. Since

$$J[y^*] = J[y^*] + \langle 0, y - y^*\rangle \,,$$

this is equivalent to $0 \in \partial J[y^*]$. $\square$

Sufficient conditions

Let X be a normed vector space and $J : X \to \mathbb{R}_\infty$ convex. $u \in X'$ is called *subgradient* of J, if

$$J[y] + \langle u, x - y\rangle \leq J[x] \text{ for all } x \in X.$$

$\partial J[y]$, the set of all subgradients of J at y, is called *subdifferential* of J.

Proposition Let X be a normed vector space and $J : X \to \mathbb{R}_\infty$ convex. Then,

$$y^* \in \operatorname*{argmin}_{y \in X} J[y] \;\Leftrightarrow\; 0 \in \partial J[y^*].$$

Proof y^* is a minimizer, if and only if $J[y^*] \leq J[y]$ for all $y \in X$. Since

$$J[y^*] = J[y^*] + \langle 0, y - y^*\rangle\,,$$

this is equivalent to $0 \in \partial J[y^*]$. □

Proposition Let $X = \mathbb{R}^d$ and $J \in C^1(X)$ convex. Then,

$$\partial J[y] = \{\nabla J[y]\} \text{ for all } y \in X.$$

Minimization using the proximal operator

Closed proper convex functionals

Let X be a vector space and $J : X \to \mathbb{R}_\infty$.

- $\operatorname{dom}(J) = \{x \in X \ : \ J[x] < \infty\}$ is called *effective domain* of J.

Closed proper convex functionals

Let X be a vector space and $J : X \to \mathbb{R}_\infty$.

- $\mathrm{dom}(J) = \{x \in X \,:\, J[x] < \infty\}$ is called *effective domain* of J.
- $\mathrm{epi}(J) = \{(x,t) \in X \times \mathbb{R} \,:\, J[x] \leq t\}$ is called *epigraph* of J.

Closed proper convex functionals

Let X be a vector space and $J : X \to \mathbb{R}_\infty$.

- $\mathrm{dom}(J) = \{x \in X \,:\, J[x] < \infty\}$ is called *effective domain* of J.
- $\mathrm{epi}(J) = \{(x,t) \in X \times \mathbb{R} \,:\, J[x] \leq t\}$ is called *epigraph* of J.
- A convex J is called *proper*, if $\mathrm{epi}(J) \neq \emptyset$,

Closed proper convex functionals

Let X be a vector space and $J : X \to \mathbb{R}_\infty$.

- $\operatorname{dom}(J) = \{x \in X \,:\, J[x] < \infty\}$ is called *effective domain* of J.
- $\operatorname{epi}(J) = \{(x,t) \in X \times \mathbb{R} \,:\, J[x] \leq t\}$ is called *epigraph* of J.
- A convex J is called *proper*, if $\operatorname{epi}(J) \neq \emptyset$,
  i.e. if there exists $x \in X$ with $J[x] < \infty$.

Closed proper convex functionals

Let X be a vector space and $J : X \to \mathbb{R}_\infty$.

- $\mathrm{dom}(J) = \{x \in X \,:\, J[x] < \infty\}$ is called *effective domain* of J.
- $\mathrm{epi}(J) = \{(x,t) \in X \times \mathbb{R} \,:\, J[x] \leq t\}$ is called *epigraph* of J.
- A convex J is called *proper*, if $\mathrm{epi}(J) \neq \emptyset$,
  i.e. if there exists $x \in X$ with $J[x] < \infty$.
- If X is a normed vector space, a proper convex functional J is called *closed*, if $\mathrm{epi}(J)$ is closed.

Closed proper convex functionals

Let X be a vector space and $J : X \to \mathbb{R}_\infty$.

- $\operatorname{dom}(J) = \{x \in X \,:\, J[x] < \infty\}$ is called *effective domain* of J.
- $\operatorname{epi}(J) = \{(x,t) \in X \times \mathbb{R} \,:\, J[x] \leq t\}$ is called *epigraph* of J.
- A convex J is called *proper*, if $\operatorname{epi}(J) \neq \emptyset$, i.e. if there exists $x \in X$ with $J[x] < \infty$.
- If X is a normed vector space, a proper convex functional J is called *closed*, if $\operatorname{epi}(J)$ is closed.
- $\Gamma_0(X)$ denotes the set of closed proper convex functionals on X.

The proximal operator

Proposition Let X be a reflexive Banach space and $J \in \Gamma_0(X)$.

Then, the mapping

$$\mathrm{prox}_J : X \to X, y \mapsto \operatorname*{argmin}_{u \in X} \left(J[u] + \frac{1}{2} \|u - y\|^2 \right)$$

is well-defined (i.e. there is a unique minimizer) and is called *proximal mapping / proximal operator*.

Proposition Let X be a reflexive Banach space and $J \in \Gamma_0(X)$.
Then, the mapping

$$\operatorname{prox}_J : X \to X, y \mapsto \operatorname*{argmin}_{u \in X} \left(J[u] + \frac{1}{2} \|u - y\|^2 \right)$$

is well-defined (i.e. there is a unique minimizer) and is called *proximal mapping / proximal operator*.

- Existence of a minimizer can be shown with the direct method.

Proposition Let X be a reflexive Banach space and $J \in \Gamma_0(X)$.
Then, the mapping

$$\operatorname{prox}_J : X \to X, y \mapsto \operatorname*{argmin}_{u \in X} \left(J[u] + \frac{1}{2} \|u - y\|^2 \right)$$

is well-defined (i.e. there is a unique minimizer) and is called *proximal mapping / proximal operator*.

- Existence of a minimizer can be shown with the direct method.
- $\Gamma_0(X)$ was constructed such that the direct method can be applied.

The proximal operator

Proposition Let X be a reflexive Banach space and $J \in \Gamma_0(X)$.
Then, the mapping

$$\operatorname{prox}_J : X \to X, y \mapsto \operatorname*{argmin}_{u \in X} \left(J[u] + \frac{1}{2} \|u - y\|^2 \right)$$

is well-defined (i.e. there is a unique minimizer) and is called *proximal mapping* / *proximal operator*.

- Existence of a minimizer can be shown with the direct method.
- $\Gamma_0(X)$ was constructed such that the direct method can be applied.
- Uniqueness of the minimizer follows from the strict convexity of $\|\cdot\|^2$.

The proximal operator

Proposition Let X be a reflexive Banach space and $J \in \Gamma_0(X)$.
Then, the mapping

$$\operatorname{prox}_J : X \to X, y \mapsto \operatorname*{argmin}_{u \in X} \left(J[u] + \frac{1}{2} \|u - y\|^2 \right)$$

is well-defined (i.e. there is a unique minimizer) and is called *proximal mapping / proximal operator*.

- Existence of a minimizer can be shown with the direct method.
- $\Gamma_0(X)$ was constructed such that the direct method can be applied.
- Uniqueness of the minimizer follows from the strict convexity of $\|\cdot\|^2$.

Note: $J \in \Gamma_0(X)$ doesn't ensure $\operatorname*{argmin}_{x \in X} J[x] \neq \emptyset$, e.g. $X = \mathbb{R}$, $J[x] = x$.

Proposition Let X be a reflexive Banach space and $J \in \Gamma_0(X)$.
Then, the mapping

$$\operatorname{prox}_J : X \to X, y \mapsto \operatorname*{argmin}_{u \in X} \left(J[u] + \frac{1}{2} \|u - y\|^2 \right)$$

is well-defined (i.e. there is a unique minimizer) and is called *proximal mapping / proximal operator*.

- Existence of a minimizer can be shown with the direct method.
- $\Gamma_0(X)$ was constructed such that the direct method can be applied.
- Uniqueness of the minimizer follows from the strict convexity of $\|\cdot\|^2$.

Note: $J \in \Gamma_0(X)$ doesn't ensure $\operatorname*{argmin}_{x \in X} J[x] \neq \emptyset$, e.g. $X = \mathbb{R}$, $J[x] = x$.

Recall: Let X be a normed vector space and $J : X \to \mathbb{R}_\infty$ convex.
$u \in X'$ is called *subgradient* of J, if

$$J[y] + \langle u, x - y \rangle \leq J[x] \text{ for all } x \in X.$$

$\partial J[y]$, the set of all subgradients of J at y, is called *subdifferential* of J.

The proximal operator - Properties

In the following, X always is a Hilbert space and we identify X' with X.

In the following, X always is a Hilbert space and we identify X' with X.

Corollary Let $J \in \Gamma_0(X)$ and $\tau > 0$. Then,

$$y^* \in \operatorname*{argmin}_{y \in X} J[y] \Leftrightarrow y^* = \operatorname{prox}_{\tau J}[y^*].$$

The proximal operator - Properties

In the following, X always is a Hilbert space and we identify X' with X.

Corollary Let $J \in \Gamma_0(X)$ and $\tau > 0$. Then,

$$y^* \in \operatorname*{argmin}_{y \in X} J[y] \Leftrightarrow y^* = \operatorname{prox}_{\tau J}[y^*].$$

Lemma Let $J \in \Gamma_0(X)$ and $\tau > 0$. Then, for $y, y^* \in X$

$$y^* = \operatorname{prox}_{\tau J}[y] \Leftrightarrow y \in y^* + \tau \partial J[y^*].$$

In the following, X always is a Hilbert space and we identify X' with X.

Corollary Let $J \in \Gamma_0(X)$ and $\tau > 0$. Then,

$$y^* \in \operatorname*{argmin}_{y \in X} J[y] \Leftrightarrow y^* = \operatorname{prox}_{\tau J}[y^*].$$

Lemma Let $J \in \Gamma_0(X)$ and $\tau > 0$. Then, for $y, y^* \in X$

$$y^* = \operatorname{prox}_{\tau J}[y] \Leftrightarrow y \in y^* + \tau \partial J[y^*].$$

Corollary Let $J \in \Gamma_0(X)$ and $\tau > 0$. Then, for $y \in X$, we have

$$\{\operatorname{prox}_{\tau J}[y]\} = (\mathrm{id} + \tau \partial J)^{-1}[y].$$

Here, for a set-valued mapping $A : X \to \mathcal{P}(Y)$, the inversion is defined by

$$A^{-1} : Y \to \mathcal{P}(X), y \mapsto A^{-1}[y] := \{z \in X \,:\, y \in A[z]\}.$$

The proximal operator - Properties

In the following, X always is a Hilbert space and we identify X' with X.

Corollary Let $J \in \Gamma_0(X)$ and $\tau > 0$. Then,

$$y^* \in \operatorname*{argmin}_{y \in X} J[y] \Leftrightarrow y^* = \operatorname{prox}_{\tau J}[y^*].$$

Lemma Let $J \in \Gamma_0(X)$ and $\tau > 0$. Then, for $y, y^* \in X$

$$y^* = \operatorname{prox}_{\tau J}[y] \Leftrightarrow y \in y^* + \tau \partial J[y^*].$$

Corollary Let $J \in \Gamma_0(X)$ and $\tau > 0$. Then, for $y \in X$, we have

$$\{\operatorname{prox}_{\tau J}[y]\} = (\operatorname{id} + \tau \partial J)^{-1}[y].$$

Here, for a set-valued mapping $A : X \to \mathcal{P}(Y)$, the inversion is defined by

$$A^{-1} : Y \to \mathcal{P}(X), y \mapsto A^{-1}[y] := \{z \in X \ : \ y \in A[z]\}.$$

Corollary Let $J \in \Gamma_0(X)$, $\tau > 0$ and $y, y^* \in X$. Then,

$$y^* \in \partial J[y] \Leftrightarrow y = \operatorname{prox}_{\tau J}[y + \tau y^*].$$

The proximal operator - Examples

Now, confine to $X = \mathbb{R}^n$ and $\|\cdot\| = \|\cdot\|_2$, i.e. *Discretize Then Optimize*.

- For $J \equiv c \in \mathbb{R}$, we have

$$\mathrm{prox}_{\tau J}(y) = \operatorname*{argmin}_{u \in \mathbb{R}^n} \left(\tau c + \frac{1}{2} \|u - y\|_2^2 \right) = y.$$

The proximal operator - Examples

Now, confine to $X = \mathbb{R}^n$ and $\|\cdot\| = \|\cdot\|_2$, i.e. *Discretize Then Optimize*.

- For $J \equiv c \in \mathbb{R}$, we have

$$\operatorname{prox}_{\tau J}(y) = \underset{u \in \mathbb{R}^n}{\operatorname{argmin}} \left(\tau c + \frac{1}{2} \|u - y\|_2^2 \right) = y.$$

- Let $g \in \mathbb{R}^n$ and $J(y) = \frac{1}{2} \sum_{i=1}^{n} (y_i - g_i)^2 = \frac{1}{2} \|y - g\|_2^2$. Then,

$$u^* := \operatorname{prox}_{\tau J}(y) = \underset{u \in \mathbb{R}^n}{\operatorname{argmin}} \left(\frac{\tau}{2} \|u - g\|_2^2 + \frac{1}{2} \|u - y\|_2^2 \right)$$

$$\Rightarrow 0 = \tau(u^* - g) + (u^* - y) \Rightarrow \operatorname{prox}_{\tau J}(y) = \frac{y + \tau g}{1 + \tau}$$

The proximal operator - Examples

Now, confine to $X = \mathbb{R}^n$ and $\|\cdot\| = \|\cdot\|_2$, i.e. *Discretize Then Optimize*.

- For $J \equiv c \in \mathbb{R}$, we have

$$\operatorname{prox}_{\tau J}(y) = \operatorname*{argmin}_{u \in \mathbb{R}^n} \left(\tau c + \frac{1}{2} \|u - y\|_2^2 \right) = y.$$

- Let $g \in \mathbb{R}^n$ and $J(y) = \frac{1}{2} \sum_{i=1}^{n} (y_i - g_i)^2 = \frac{1}{2} \|y - g\|_2^2$. Then,

$$u^* := \operatorname{prox}_{\tau J}(y) = \operatorname*{argmin}_{u \in \mathbb{R}^n} \left(\frac{\tau}{2} \|u - g\|_2^2 + \frac{1}{2} \|u - y\|_2^2 \right)$$

$$\Rightarrow 0 = \tau(u^* - g) + (u^* - y) \Rightarrow \operatorname{prox}_{\tau J}(y) = \frac{y + \tau g}{1 + \tau}$$

- For $J(y) = \sum_{i=1}^{n} J_i(y_i)$ with $J_i \in \Gamma_0(\mathbb{R})$, we have

$$\operatorname{prox}_{\tau J}(y) = (\operatorname{prox}_{\tau J_1}(y_1), \ldots, \operatorname{prox}_{\tau J_n}(y_n)).$$

The proximal operator - Examples (cont.)

- For $J(y) := \|y\|_1$, $\operatorname{prox}_{\tau J}(y)$ is the *soft threshold* operator, i.e.

$$(\operatorname{prox}_{\tau J}(y))_i = \begin{cases} y_i - \tau & y_i \geq \tau \\ 0 & |y_i| < \tau \\ y_i + \tau & y_i \leq -\tau \end{cases} .$$

The proximal operator - Examples (cont.)

- For $J(y) := \|y\|_1$, $\mathrm{prox}_{\tau J}(y)$ is the *soft threshold* operator, i.e.

$$(\mathrm{prox}_{\tau J}(y))_i = \begin{cases} y_i - \tau & y_i \geq \tau \\ 0 & |y_i| < \tau \\ y_i + \tau & y_i \leq -\tau \end{cases}.$$

- If $C \subset \mathbb{R}^n$ is a nonempty, closed, convex set, then

$$I_C : \mathbb{R}^n \to \mathbb{R}_\infty, y \mapsto \begin{cases} 0 & y \in C \\ \infty & y \notin C, \end{cases}$$

is called *indicator function* of C in $\Gamma_0(\mathbb{R}^n)$

The proximal operator - Examples (cont.)

- For $J(y) := \|y\|_1$, $\operatorname{prox}_{\tau J}(y)$ is the *soft threshold* operator, i.e.

$$(\operatorname{prox}_{\tau J}(y))_i = \begin{cases} y_i - \tau & y_i \geq \tau \\ 0 & |y_i| < \tau \\ y_i + \tau & y_i \leq -\tau \end{cases}.$$

- If $C \subset \mathbb{R}^n$ is a nonempty, closed, convex set, then

$$I_C : \mathbb{R}^n \to \mathbb{R}_\infty, y \mapsto \begin{cases} 0 & y \in C \\ \infty & y \notin C, \end{cases}$$

  is called *indicator function* of C in $\Gamma_0(\mathbb{R}^n)$ and we have

$$\operatorname{prox}_{\tau I_C}(y) = \Pi_C(y),$$

  where Π_C is the Euclidean projection to C, i.e.

$$\Pi_C(y) = \operatorname*{argmin}_{z \in C} \|z - y\|_2 .$$

The proximal point algorithm

Finding a minimizer of $J \in \Gamma_0$ and a fixed point of $\mathrm{prox}_{\tau J}$ are equivalent:

Finding a minimizer of $J \in \Gamma_0$ and a fixed point of $\operatorname{prox}_{\tau J}$ are equivalent:

Proximal point algorithm

$$y^{k+1} = \operatorname{prox}_{\tau J}(y^k)$$

for a step size $\tau > 0$ and an initial value value $y^0 \in \mathbb{R}^n$.

The proximal point algorithm

Finding a minimizer of $J \in \Gamma_0$ and a fixed point of $\operatorname{prox}_{\tau J}$ are equivalent:

Proximal point algorithm

$$y^{k+1} = \operatorname{prox}_{\tau J}(y^k)$$

for a step size $\tau > 0$ and an initial value value $y^0 \in \mathbb{R}^n$.

If a minimizer of J exists, y^k converges to the set of minimizers and $J(y^k)$ to the optimal value.

The proximal point algorithm

Finding a minimizer of $J \in \Gamma_0$ and a fixed point of $\mathrm{prox}_{\tau J}$ are equivalent:

Proximal point algorithm

$$y^{k+1} = \mathrm{prox}_{\tau J}(y^k)$$

for a step size $\tau > 0$ and an initial value value $y^0 \in \mathbb{R}^n$.

If a minimizer of J exists, y^k converges to the set of minimizers and $J(y^k)$ to the optimal value.

If $J \in C^1(\mathbb{R}^n) \cap \Gamma_0(\mathbb{R}^n)$, then $y^{k+1} = \mathrm{prox}_{\tau J}(y^k)$ is characterized by

$$0 = \tau \nabla J(y^{k+1}) + (y^{k+1} - y^k) \;\Rightarrow\; y^{k+1} = y^k - \tau \nabla J(y^{k+1})$$

The proximal point algorithm

Finding a minimizer of $J \in \Gamma_0$ and a fixed point of $\operatorname{prox}_{\tau J}$ are equivalent:

Proximal point algorithm

$$y^{k+1} = \operatorname{prox}_{\tau J}(y^k)$$

for a step size $\tau > 0$ and an initial value value $y^0 \in \mathbb{R}^n$.

If a minimizer of J exists, y^k converges to the set of minimizers and $J(y^k)$ to the optimal value.

If $J \in C^1(\mathbb{R}^n) \cap \Gamma_0(\mathbb{R}^n)$, then $y^{k+1} = \operatorname{prox}_{\tau J}(y^k)$ is characterized by

$$0 = \tau \nabla J(y^{k+1}) + (y^{k+1} - y^k) \;\Rightarrow\; y^{k+1} = y^k - \tau \nabla J(y^{k+1})$$

This is the backward Euler discretization of the gradient descent of J.

The proximal point algorithm

Finding a minimizer of $J \in \Gamma_0$ and a fixed point of $\operatorname{prox}_{\tau J}$ are equivalent:

Proximal point algorithm

$$y^{k+1} = \operatorname{prox}_{\tau J}(y^k)$$

for a step size $\tau > 0$ and an initial value value $y^0 \in \mathbb{R}^n$.

If a minimizer of J exists, y^k converges to the set of minimizers and $J(y^k)$ to the optimal value.

If $J \in C^1(\mathbb{R}^n) \cap \Gamma_0(\mathbb{R}^n)$, then $y^{k+1} = \operatorname{prox}_{\tau J}(y^k)$ is characterized by

$$0 = \tau \nabla J(y^{k+1}) + (y^{k+1} - y^k) \;\Rightarrow\; y^{k+1} = y^k - \tau \nabla J(y^{k+1})$$

This is the backward Euler discretization of the gradient descent of J.

Thus, for differentiable J, the proximal point algorithm is equivalent to the fully implicit gradient descent.

The proximal point algorithm

Finding a minimizer of $J \in \Gamma_0$ and a fixed point of $\operatorname{prox}_{\tau J}$ are equivalent:

Proximal point algorithm

$$y^{k+1} = \operatorname{prox}_{\tau J}(y^k)$$

for a step size $\tau > 0$ and an initial value value $y^0 \in \mathbb{R}^n$.

If a minimizer of J exists, y^k converges to the set of minimizers and $J(y^k)$ to the optimal value.

If $J \in C^1(\mathbb{R}^n) \cap \Gamma_0(\mathbb{R}^n)$, then $y^{k+1} = \operatorname{prox}_{\tau J}(y^k)$ is characterized by

$$0 = \tau \nabla J(y^{k+1}) + (y^{k+1} - y^k) \;\Rightarrow\; y^{k+1} = y^k - \tau \nabla J(y^{k+1})$$

This is the backward Euler discretization of the gradient descent of J.

Thus, for differentiable J, the proximal point algorithm is equivalent to the fully implicit gradient descent.

Note: For most image processing problems, this algorithm is not very practical, since $\operatorname{prox}_{\tau J}$ can't be evaluated efficiently.

The proximal point algorithm

Finding a minimizer of $J \in \Gamma_0$ and a fixed point of $\operatorname{prox}_{\tau J}$ are equivalent:

Proximal point algorithm

$$y^{k+1} = \operatorname{prox}_{\tau J}(y^k)$$

for a step size $\tau > 0$ and an initial value value $y^0 \in \mathbb{R}^n$.

If a minimizer of J exists, y^k converges to the set of minimizers and $J(y^k)$ to the optimal value.

If $J \in C^1(\mathbb{R}^n) \cap \Gamma_0(\mathbb{R}^n)$, then $y^{k+1} = \operatorname{prox}_{\tau J}(y^k)$ is characterized by

$$0 = \tau \nabla J(y^{k+1}) + (y^{k+1} - y^k) \;\Rightarrow\; y^{k+1} = y^k - \tau \nabla J(y^{k+1})$$

This is the backward Euler discretization of the gradient descent of J.

Thus, for differentiable J, the proximal point algorithm is equivalent to the fully implicit gradient descent.

Note: For most image processing problems, this algorithm is not very practical, since $\operatorname{prox}_{\tau J}$ can't be evaluated efficiently.

$\Rightarrow$ Operator Splitting.

For $J = G + H$, we consider the optimization problem

$$\min_{y \in \mathbb{R}^n} (G(y) + H(y)),$$

where $G \in C^1(\mathbb{R}^n) \cap \Gamma_0(\mathbb{R}^n)$ and $H \in \Gamma_0(\mathbb{R}^n)$.

The proximal gradient algorithm

For $J = G + H$, we consider the optimization problem

$$\min_{y \in \mathbb{R}^n} (G(y) + H(y)),$$

where $G \in C^1(\mathbb{R}^n) \cap \Gamma_0(\mathbb{R}^n)$ and $H \in \Gamma_0(\mathbb{R}^n)$.

Proximal gradient algorithm

$$y^{k+1} = \operatorname{prox}_{\tau_k H}\left(y^k - \tau_k \nabla G(y^k)\right)$$

for step sizes $\tau_k > 0$ and an initial value $y^0 \in \mathbb{R}^n$.

The proximal gradient algorithm

For $J = G + H$, we consider the optimization problem

$$\min_{y\in\mathbb{R}^n} (G(y) + H(y)),$$

where $G \in C^1(\mathbb{R}^n) \cap \Gamma_0(\mathbb{R}^n)$ and $H \in \Gamma_0(\mathbb{R}^n)$.

Proximal gradient algorithm

$$y^{k+1} = \mathrm{prox}_{\tau_k H}\left(y^k - \tau_k \nabla G(y^k)\right)$$

for step sizes $\tau_k > 0$ and an initial value $y^0 \in \mathbb{R}^n$. Using

$$\mathcal{F}_\tau(y) = \frac{1}{\tau}\left(y - \mathrm{prox}_{\tau H}\left(y - \tau\nabla G(y)\right)\right),$$

we get

$$y^{k+1} = y^k - \tau_k \mathcal{F}_{\tau_k}(y^k).$$

The proximal gradient algorithm

For $J = G + H$, we consider the optimization problem

$$\min_{y \in \mathbb{R}^n} (G(y) + H(y)),$$

where $G \in C^1(\mathbb{R}^n) \cap \Gamma_0(\mathbb{R}^n)$ and $H \in \Gamma_0(\mathbb{R}^n)$.

Proximal gradient algorithm

$$y^{k+1} = \operatorname{prox}_{\tau_k H}\left(y^k - \tau_k \nabla G(y^k)\right)$$

for step sizes $\tau_k > 0$ and an initial value $y^0 \in \mathbb{R}^n$. Using

$$\mathcal{F}_\tau(y) = \frac{1}{\tau}\left(y - \operatorname{prox}_{\tau H}\left(y - \tau \nabla G(y)\right)\right),$$

we get

$$y^{k+1} = y^k - \tau_k \mathcal{F}_{\tau_k}(y^k).$$

This is also called *forward-backward splitting*, since it combines

- a forward Euler gradient descent step in G with
- a proximal point algorithm step in H (equivalent to a backward Euler gradient descent step in H).

The proximal gradient algorithm - Properties

Assumptions as before and ∇G Lipschitz continuous with constant $L > 0$.

The proximal gradient algorithm - Properties

Assumptions as before and ∇G Lipschitz continuous with constant $L > 0$.

Lemma For all $y, z \in \mathbb{R}^n$ and $\tau \in [0, \frac{1}{L}]$, it holds that

$$J(y - \tau \mathcal{F}_\tau(y)) \leq J(z) + \mathcal{F}_\tau(y) \cdot (y - z) - \frac{\tau}{2} \left\| \mathcal{F}_\tau(y) \right\|_2^2 .$$

The proximal gradient algorithm - Properties

Assumptions as before and ∇G Lipschitz continuous with constant $L > 0$.

Lemma For all $y, z \in \mathbb{R}^n$ and $\tau \in [0, \frac{1}{L}]$, it holds that

$$J(y - \tau \mathcal{F}_\tau(y)) \leq J(z) + \mathcal{F}_\tau(y) \cdot (y - z) - \frac{\tau}{2} \left\| \mathcal{F}_\tau(y) \right\|_2^2 .$$

Theorem Let $\tau_k \in [\tau_{\text{min}}, \frac{1}{L}]$, where $\tau_{\text{min}} \in (0, \frac{1}{L}]$, and let $\operatorname*{argmin}_{x \in X} J \neq \emptyset$. Then, the proximal gradient algorithm converges. More precisely,

$$0 \leq J(y^k) - J(y^*) \leq \frac{1}{2k\tau_{\text{min}}} \left\| y^0 - y^* \right\|_2^2 = O\left(\tfrac{1}{k}\right)$$

for $y^* \in \operatorname*{argmin}_{x \in X} J$. Moreover, $(y^k)_k$ converges to the set of minimizers, i.e.

$$\lim_{k \to \infty} \operatorname{dist}(y^k, \operatorname{argmin} J) = 0.$$

The proximal gradient algorithm - Special cases

The above also proves convergence of other methods.

The proximal gradient algorithm - Special cases

The above also proves convergence of other methods.

- $G = 0$, $H = J \Rightarrow$ proximal point algorithm
  Since $\nabla 0$ is Lipschitz continuous with constant 0, it follows the convergence for arbitrary, bounded step sizes.

The proximal gradient algorithm - Special cases

The above also proves convergence of other methods.

- $G = 0$, $H = J \Rightarrow$ proximal point algorithm
  Since $\nabla 0$ is Lipschitz continuous with constant 0, it follows the convergence for arbitrary, bounded step sizes.
- $G = J$, $H = 0 \Rightarrow$ fully explicit gradient descent
  If ∇J is Lipschitz continuous, we get convergence for suitable τ_n.

The proximal gradient algorithm - Special cases

The above also proves convergence of other methods.

- $G = 0$, $H = J \Rightarrow$ proximal point algorithm
  Since $\nabla 0$ is Lipschitz continuous with constant 0, it follows the convergence for arbitrary, bounded step sizes.
- $G = J$, $H = 0 \Rightarrow$ fully explicit gradient descent
  If ∇J is Lipschitz continuous, we get convergence for suitable τ_n.
- $C \subset \mathbb{R}^n$ nonempty, convex and closed, $G = J$, $H = I_C$
  $\Rightarrow$ *projected gradient descent*, which minimizes $J(y)$ under the constraint $y \in C$. If ∇J is Lipschitz continuous, we get convergence for suitable τ_n.

Primal-dual approaches

Let $J : X \to \mathbb{R}_\infty$ be proper. Then, *Fenchel conjugate* of J denotes

$$J^* : X' :\to \mathbb{R}_\infty, x' \mapsto \sup_{x \in X} \left(\langle x', x \rangle - J[x] \right)$$

Primal-dual approaches

Let $J : X \to \mathbb{R}_\infty$ be proper. Then, *Fenchel conjugate* of J denotes

$$J^* : X' :\to \mathbb{R}_\infty, x' \mapsto \sup_{x \in X} \left(\langle x', x \rangle - J[x] \right)$$

Particularly relevant in image processing are problems of the type

$$\min_{y \in \mathbb{R}^n} (G(y) + H(Ay)),$$

where $G \in \Gamma_0(\mathbb{R}^n)$, $H \in \Gamma_0(\mathbb{R}^m)$ and $A : \mathbb{R}^n \to \mathbb{R}^m$ is linear.

Primal-dual approaches

Let $J : X \to \mathbb{R}_\infty$ be proper. Then, *Fenchel conjugate* of J denotes

$$J^* : X' :\to \mathbb{R}_\infty, x' \mapsto \sup_{x \in X} \left(\langle x', x \rangle - J[x] \right)$$

Particularly relevant in image processing are problems of the type

$$\min_{y \in \mathbb{R}^n} (G(y) + H(Ay)),$$

where $G \in \Gamma_0(\mathbb{R}^n)$, $H \in \Gamma_0(\mathbb{R}^m)$ and $A : \mathbb{R}^n \to \mathbb{R}^m$ is linear. Then,

$$\begin{aligned} \inf_{y \in \mathbb{R}^n} (G(y) + H(Ay)) &\overset{H = H^{**}}{=} \inf_{y \in \mathbb{R}^n} (G(y) + H^{**}(Ay)) \\ &= \inf_{y \in \mathbb{R}^n} (G(y) + \sup_{z \in \mathbb{R}^m} (Ay \cdot z - H^*(z))) \\ &= \inf_{y \in \mathbb{R}^n} \sup_{z \in \mathbb{R}^m} (Ay \cdot z + G(y) - H^*(z)) . \end{aligned}$$

Primal-dual approaches

Let $J : X \to \mathbb{R}_\infty$ be proper. Then, *Fenchel conjugate* of J denotes

$$J^* : X' :\to \mathbb{R}_\infty, x' \mapsto \sup_{x \in X} \left(\langle x', x \rangle - J[x] \right)$$

Particularly relevant in image processing are problems of the type

$$\min_{y \in \mathbb{R}^n} (G(y) + H(Ay)),$$

where $G \in \Gamma_0(\mathbb{R}^n)$, $H \in \Gamma_0(\mathbb{R}^m)$ and $A : \mathbb{R}^n \to \mathbb{R}^m$ is linear. Then,

$$\begin{aligned} \inf_{y \in \mathbb{R}^n} (G(y) + H(Ay)) &\overset{H=H^{**}}{=} \inf_{y \in \mathbb{R}^n} (G(y) + H^{**}(Ay)) \\ &= \inf_{y \in \mathbb{R}^n} (G(y) + \sup_{z \in \mathbb{R}^m} (Ay \cdot z - H^*(z))) \\ &= \inf_{y \in \mathbb{R}^n} \sup_{z \in \mathbb{R}^m} (Ay \cdot z + G(y) - H^*(z)) \,. \end{aligned}$$

The necessary conditions for z and y are

$$0 \in \partial_z (Ay \cdot z - H^*(z)) = Ay - \partial H^*(z) \ \Rightarrow \ Ay \in \partial H^*(z),$$

$$0 \in \partial_y \left(A^T z \cdot y + G(y) \right) = A^T z + \partial G(y) \ \Rightarrow \ -A^T z \in \partial G(y).$$

Primal-dual approaches (cont.)

$Ay \in \partial H^*(z) \;\wedge\; -A^T z \in \partial G(y)$ is equivalent to

$$z = \mathrm{prox}_{\sigma H^*}[z + \sigma A y] \;\wedge\; y = \mathrm{prox}_{\tau G}[y - \tau A^T z]$$

for $\tau, \sigma > 0$

Primal-dual approaches (cont.)

$Ay \in \partial H^*(z) \ \wedge \ -A^T z \in \partial G(y)$ is equivalent to

$$z = \text{prox}_{\sigma H^*}[z + \sigma A y] \ \wedge \ y = \text{prox}_{\tau G}[y - \tau A^T z]$$

for $\tau, \sigma > 0$ and motivates the algorithm

$$\begin{aligned} z^{k+1} &= \text{prox}_{\sigma H^*}[z^k + \sigma A \overline{y}^k] \\ y^{k+1} &= \text{prox}_{\tau G}[y^k - \tau A^T z^{k+1}] \\ \overline{y}^{k+1} &= y^{k+1} + \theta(y^{k+1} - y^k) \qquad \text{(extrapolation step)} \end{aligned}$$

for $\theta \in [0, 1]$, $\overline{y}^0 = y^0 \in \mathbb{R}^n$, $z^0 \in \mathbb{R}^m$.

Primal-dual approaches (cont.)

$Ay \in \partial H^*(z) \;\wedge\; -A^T z \in \partial G(y)$ is equivalent to

$$z = \operatorname{prox}_{\sigma H^*}[z + \sigma A y] \;\wedge\; y = \operatorname{prox}_{\tau G}[y - \tau A^T z]$$

for $\tau, \sigma > 0$ and motivates the algorithm

$$\begin{aligned} z^{k+1} &= \operatorname{prox}_{\sigma H^*}[z^k + \sigma A \overline{y}^k] \\ y^{k+1} &= \operatorname{prox}_{\tau G}[y^k - \tau A^T z^{k+1}] \\ \overline{y}^{k+1} &= y^{k+1} + \theta(y^{k+1} - y^k) \qquad \text{(extrapolation step)} \end{aligned}$$

for $\theta \in [0,1]$, $\overline{y}^0 = y^0 \in \mathbb{R}^n$, $z^0 \in \mathbb{R}^m$.

- avoids computing $\operatorname{prox}_{\sigma H(A\cdot)}$, just needs $\operatorname{prox}_{\sigma H^*}$, A, A^T
- typically $\operatorname{prox}_{\sigma H^*}$ can be computed pointwise
- proposed in [Chambolle, Pock '10], very popular ($4300+$ citations)
- also called *Primal-Dual Hybrid Gradient* (PDHG) method
- well suited for models that use the total variation as regularizer
- convergence guaranteed if $\theta = 1$, $\tau\sigma \|A\|^2 < 1$

PDHG extensions and variants

- $\theta = 0$: Arrow–Hurwicz algorithm, used for TV minimization before [Zhu, Chan '08]
- diagonal preconditioning [Pock, Chambolle '11]
- extension to Banach spaces with applications to inverse problems [Hohage, Homann '14]
- stochastic extension based on arbitrary sampling of the dual variables [Chambolle, Ehrhardt, Richtárik, Schönlieb, '18]
- stochastic PDHG to solve regularized stochastic minimization problems [Qiao, Lin, Qin, Lu Neurocomputing '18]
- learned primal-dual reconstruction [Adler, Öktem TMI '18]
- Riemannian Chambolle–Pock algorithm [Bergmann, Herzog, Louzeiro, Tenbrinck Vidal-Núñez '21]
- ...

A primal-dual approach for the ROF model

ROF model $J[y] = \int_\Omega (y - f)^2 \, \mathrm{d}x + \lambda \int_\Omega \|\nabla y(x)\| \, \mathrm{d}x$

A primal-dual approach for the ROF model

ROF model $J[y] = \int_\Omega (y - f)^2 \, \mathrm{d}x + \lambda \int_\Omega \|\nabla y(x)\| \, \mathrm{d}x$

Discretization: $y = (y_{i,j}) \in X := \mathbb{R}^{M \times N}$, grid width h.

A primal-dual approach for the ROF model

$$\text{ROF model } J[y] = \int_\Omega (y - f)^2 \,\mathrm{d}x + \lambda \int_\Omega \|\nabla y(x)\| \,\mathrm{d}x$$

Discretization: $y = (y_{i,j}) \in X := \mathbb{R}^{M\times N}$, grid width h.

$$G(y) = \frac{1}{2\lambda} \|y - f\|_2^2 \;\Rightarrow\; \mathrm{prox}_{\tau G}(y) = \frac{y + \frac{\tau}{\lambda} f}{1 + \frac{\tau}{\lambda}}.$$

A primal-dual approach for the ROF model

$$\text{ROF model } J[y] = \int_\Omega (y-f)^2\,\mathrm{d}x + \lambda \int_\Omega \|\nabla y(x)\|\,\mathrm{d}x$$

Discretization: $y = (y_{i,j}) \in X := \mathbb{R}^{M\times N}$, grid width h.

$$G(y) = \frac{1}{2\lambda}\|y-f\|_2^2 \;\Rightarrow\; \mathrm{prox}_{\tau G}(y) = \frac{y + \frac{\tau}{\lambda} f}{1+\frac{\tau}{\lambda}}.$$

Forward differences: $(\nabla^h y)_{i,j} = ((\partial_1^{h+} y)_{i,j}, (\partial_2^{h+} y)_{i,j}) \in X \times X$, where

$$(\partial_1^{h+} y)_{i,j} = \begin{cases} \frac{y_{i+1,j}-y_{i,j}}{h} & i < M; \\ 0 & i = M; \end{cases}, \; j = 1,\dots,N,$$

$$(\partial_2^{h+} y)_{i,j} = \begin{cases} \frac{y_{i,j+1}-y_{i,j}}{h} & j < N; \\ 0 & j = N; \end{cases}, \; i = 1,\dots,M.$$

$$H(\nabla^h y) := ||\nabla^h y||_1 := \sum_{i,j} |(\nabla^h y)_{i,j}|.$$

A primal-dual approach for the ROF model

$$\text{ROF model } J[y] = \int_\Omega (y-f)^2 \,\mathrm{d}x + \lambda \int_\Omega \|\nabla y(x)\| \,\mathrm{d}x$$

Discretization: $y = (y_{i,j}) \in X := \mathbb{R}^{M \times N}$, grid width h.

$$G(y) = \frac{1}{2\lambda} \|y - f\|_2^2 \;\Rightarrow\; \operatorname{prox}_{\tau G}(y) = \frac{y + \frac{\tau}{\lambda} f}{1 + \frac{\tau}{\lambda}}.$$

Forward differences: $(\nabla^h y)_{i,j} = ((\partial_1^{h+} y)_{i,j}, (\partial_2^{h+} y)_{i,j}) \in X \times X$, where

$$(\partial_1^{h+} y)_{i,j} = \begin{cases} \frac{y_{i+1,j} - y_{i,j}}{h} & i < M; \\ 0 & i = M; \end{cases}, \quad j = 1, \ldots, N,$$

$$(\partial_2^{h+} y)_{i,j} = \begin{cases} \frac{y_{i,j+1} - y_{i,j}}{h} & j < N; \\ 0 & j = N; \end{cases}, \quad i = 1, \ldots, M.$$

$$H(\nabla^h y) := ||\nabla^h y||_1 := \sum_{i,j} |(\nabla^h y)_{i,j}|.$$

With $\mathbb{R}^n \simeq X$, $\mathbb{R}^m \simeq X \times X$ and $A \simeq \nabla^h$, $G(y) + H(Ay)$ is a discretization of the ROF-model fitting the framework.

A primal-dual approach for the ROF model (cont.)

One can show, that $H^* = I_P$, where

$$P = \left\{ p \in X \times X \;:\; \|p\|_\infty := \max_{i,j} |p_{i,j}| \leq 1 \right\}.$$

Thus, $\text{prox}_{\sigma H^*} = \Pi_P$.

A primal-dual approach for the ROF model (cont.)

One can show, that $H^* = I_P$, where

$$P = \left\{ p \in X \times X \;:\; \|p\|_\infty := \max_{i,j} |p_{i,j}| \leq 1 \right\}.$$

Thus, $\operatorname{prox}_{\sigma H^*} = \Pi_P$. Moreover, for $p \in X \times X$, we have

$$(\Pi_P(p))_{i,j} = \frac{p_{i,j}}{\max(1, |p_{i,j}|)}.$$

A primal-dual approach for the ROF model (cont.)

One can show, that $H^* = I_P$, where

$$P = \left\{ p \in X \times X \; : \; \|p\|_\infty := \max_{i,j} |p_{i,j}| \leq 1 \right\}.$$

Thus, $\mathrm{prox}_{\sigma H^*} = \Pi_P$. Moreover, for $p \in X \times X$, we have

$$(\Pi_P(p))_{i,j} = \frac{p_{i,j}}{\max(1, |p_{i,j}|)}.$$

Thus, the full algorithm is

$$\begin{aligned} z^{k+1} &= \Pi_P(z^k + \sigma A \overline{y}^k) \\ y^{k+1} &= \frac{y^k - \tau A^T z^{k+1} + \frac{\tau}{\lambda} f}{1 + \frac{\tau}{\lambda}} \\ \overline{y}^{k+1} &= y^{k+1} + \theta(y^{k+1} - y^k) \end{aligned}$$

A primal-dual approach for the ROF model (cont.)

One can show, that $H^* = I_P$, where

$$P = \left\{ p \in X \times X \;:\; \|p\|_\infty := \max_{i,j} |p_{i,j}| \leq 1 \right\}.$$

Thus, $\operatorname{prox}_{\sigma H^*} = \Pi_P$. Moreover, for $p \in X \times X$, we have

$$(\Pi_P(p))_{i,j} = \frac{p_{i,j}}{\max(1, |p_{i,j}|)}.$$

Thus, the full algorithm is

$$\begin{aligned} z^{k+1} &= \Pi_P(z^k + \sigma A \overline{y}^k) \\ y^{k+1} &= \frac{y^k - \tau A^T z^{k+1} + \frac{\tau}{\lambda} f}{1 + \frac{\tau}{\lambda}} \\ \overline{y}^{k+1} &= y^{k+1} + \theta(y^{k+1} - y^k) \end{aligned}$$

Moreover, one can show $\left|\left|\left|\nabla^h\right|\right|\right|^2 \leq \frac{8}{h^2}$.

A primal-dual approach for the ROF model (cont.)

One can show, that $H^* = I_P$, where

$$P = \left\{ p \in X \times X \ : \ \|p\|_\infty := \max_{i,j} |p_{i,j}| \leq 1 \right\}.$$

Thus, $\operatorname{prox}_{\sigma H^*} = \Pi_P$. Moreover, for $p \in X \times X$, we have

$$(\Pi_P(p))_{i,j} = \frac{p_{i,j}}{\max(1, |p_{i,j}|)}.$$

Thus, the full algorithm is

$$\begin{aligned} z^{k+1} &= \Pi_P(z^k + \sigma A \overline{y}^k) \\ y^{k+1} &= \frac{y^k - \tau A^T z^{k+1} + \frac{\tau}{\lambda} f}{1 + \frac{\tau}{\lambda}} \\ \overline{y}^{k+1} &= y^{k+1} + \theta(y^{k+1} - y^k) \end{aligned}$$

Moreover, one can show $\left\|\left|\nabla^h\right|\right\|^2 \leq \frac{8}{h^2}$.

`https://berkels.github.io/jupyterlite-demo`

A primal-dual approach for the ROF model (cont.)

One can show, that $H^* = I_P$, where

$$P = \left\{ p \in X \times X \,:\, \|p\|_\infty := \max_{i,j} |p_{i,j}| \leq 1 \right\}.$$

Thus, $\operatorname{prox}_{\sigma H^*} = \Pi_P$. Moreover, for $p \in X \times X$, we have

$$(\Pi_P(p))_{i,j} = \frac{p_{i,j}}{\max(1, |p_{i,j}|)}.$$

Thus, the full algorithm is

$$\begin{aligned}
z^{k+1} &= \Pi_P(z^k + \sigma A \overline{y}^k) \\
y^{k+1} &= \frac{y^k - \tau A^T z^{k+1} + \frac{\tau}{\lambda} f}{1 + \frac{\tau}{\lambda}} \\
\overline{y}^{k+1} &= y^{k+1} + \theta(y^{k+1} - y^k)
\end{aligned}$$

`https://berkels.github.io/`

`jupyterlite-demo`

Moreover, one can show $\left|\left|\left|\nabla^h\right|\right|\right|^2 \leq \frac{8}{h^2}$.

There are variants of the primal-dual method, which exploit the strict convexity of the data term G for an even faster convergence.

Mumford-Shah based image segmentation