Chapter 4
Variational Image Restoration

Computer Vision I: Variational Methods

Winter 2017/18

Dr. Yvain QUÉAU
Chair for Computer Vision and Pattern Recognition
Departments of Informatics & Mathematics
Technical University of Munich

Overview

1. Inverse Problems and Image Restoration
2. Image Denoising
3. Image Deblurring
4. Inverse Problems and Bayesian Inference
5. Motion Blur and Defocus Blur
6. Video Super Resolution
7. Inpainting
8. Numerical Solving

Overview

1. **Inverse Problems and Image Restoration**
2. Image Denoising
3. Image Deblurring
4. Inverse Problems and Bayesian Inference
5. Motion Blur and Defocus Blur
6. Video Super Resolution
7. Inpainting
8. Numerical Solving

Dr. Yvain Quéau

Inverse Problems, Ill-Posedness and Regularization

In mathematics, the conversion of measurement data into information about the observed object or the observed physical system is referred to as an inverse problem.

Following Hadamard (1902), a mathematical problem is called well-posed iff:

1. A solution exists.
2. The solution is unique.
3. The solution's behavior changes continuously with the initial conditions.

Inverse problems are often ill-posed. Since the measurement data is often not sufficient to uniquely characterize the observed object or system, one introduces prior knowledge to disambiguate which solutions are apriori more likely. In the context of variational methods this prior knowledge gives rise to the regularity term.

Overview

1. Inverse Problems and Image Restoration
2. **Image Denoising**
3. Image Deblurring
4. Inverse Problems and Bayesian Inference
5. Motion Blur and Defocus Blur
6. Video Super Resolution
7. Inpainting
8. Numerical Solving

Image Restoration: Denoising

Image restoration is a classical inverse problem: Given an observed image $f : \Omega \to \mathbb{R}$ and a (typically stochastic) model of an image degradation process, we want to restore the original image $u : \Omega \to \mathbb{R}$.

Image denoising is an example of image restoration where we assume that the true image u is corrupted by (additive) noise:

$$f = u + \eta, \qquad \eta \sim \mathcal{N}(0, \sigma).$$

Rudin, Osher, Fatemi (1992) denoise f by minimizing a quadratic data term with Total Variation (TV) regularization:

$$\min_u \frac{1}{2} \int |u - f|^2 dx + \int |\nabla u| dx.$$

This gives rise to the Euler-Lagrange equation

$$u - f - \operatorname{div}\left(\frac{\nabla u}{|\nabla u|}\right) = 0.$$

Other noise models and regularizers are conceivable.

Image Restoration: Denoising

original noisy denoised

(Goldlücke, Strekalovskiy, Cremers, SIAM J. Imaging Sci. ’12)

Overview

1. Inverse Problems and Image Restoration
2. Image Denoising
3. **Image Deblurring**
4. Inverse Problems and Bayesian Inference
5. Motion Blur and Defocus Blur
6. Video Super Resolution
7. Inpainting
8. Numerical Solving

Image Restoration: Deblurring

A prototypical blur model is given by

$$f = A * u + \eta \qquad \eta \sim \mathcal{N}(0, \sigma),$$

with a blur kernel A.

In a variational setting, this process can be inverted by minimizing the TV deblurring functional:

$$\min_u \frac{1}{2} \int |A * u - f|^2 dx + \int |\nabla u| dx.$$

For symmetric kernels A, the Euler-Lagrange equation is given by:

$$A * (A * u - f) - \operatorname{div}\left(\frac{\nabla u}{|\nabla u|}\right) = 0,$$

and the gradient descent equation

$$\frac{\partial u}{\partial t} = -A * (A * u - f) + \operatorname{div}\left(\frac{\nabla u}{|\nabla u|}\right).$$

Image Restoration: Deblurring

Original blurred and noisy deblurred

(Goldluecke, Cremers, ICCV 2011)

Overview

Inverse Problems and Bayesian Inference

The framework of Bayesian inference allows to systematically derive functionals for different image formation models.

Let u be the unknown true image and f the observed one, then we can write the joint probability for u and f as:

$$\mathcal{P}(u,f) = \mathcal{P}(u|f)\,\mathcal{P}(f) = \mathcal{P}(f|u)\mathcal{P}(u).$$

Rewriting this expression we obtain the Bayesian formula (Thomas Bayes 1887):

$$\mathcal{P}(u|f) = \frac{\mathcal{P}(f|u)\,\mathcal{P}(u)}{\mathcal{P}(f)}.$$

Maximum Aposteriori (MAP) estimation aims at computing the most likely solution $\hat{u}$ given f by maximizing the posterior probability $\mathcal{P}(u|f)$

$$\hat{u} = \arg\max_u \mathcal{P}(u|f) = \arg\max_u \mathcal{P}(f|u)\,\mathcal{P}(u).$$

$\mathcal{P}(f|u)$ is called the likelihood and $\mathcal{P}(u)$ the prior.

MAP Estimation in the Discrete Setting

Let us assume n independent pixels. For each the measured intensity f_i is given by the true intensity u_i plus additive Gaussian noise. This corresponds to the likelihood

$$\mathcal{P}(f_i|u_i) \propto \exp\left(-\frac{(u_i - f_i)^2}{2\sigma^2}\right).$$

Since all measurements are mutually independent, we obtain for the entire vector $f = (f_1, \ldots, f_n)$ of pixel intensities:

$$\mathcal{P}(f|u) = \prod_{i=1}^{n} \mathcal{P}(f_i|u) = \prod_{i=1}^{n} \mathcal{P}(f_i|u_i) \propto \prod_{i=1}^{n} \exp\left(-\frac{(u_i - f_i)^2}{2\sigma^2}\right).$$

We now expand the prior:

$$\mathcal{P}(u) = \mathcal{P}(u_1 \ldots u_n) = \mathcal{P}(u_1|u_2 \ldots u_n)\mathcal{P}(u_2 \ldots u_n) \propto \prod_{i=1}^{n-1} \mathcal{P}(u_i|u_{i+1}),$$

where we assumed a Markov property, namely that the probability of u_i is sufficiently characterized by its neighbor.

MAP Estimation in the Discrete Setting

Assuming a simple smoothness prior, we have:

$$\mathcal{P}(u) \propto \prod_{i=1}^{n-1} \mathcal{P}(u_i|u_{i+1}) \propto \prod_{i=1}^{n-1} \exp\left(-\lambda|u_i - u_{i+1}|\right).$$

With these assumptions, the posterior distribution is given by:

$$\mathcal{P}(u|f) \propto \prod_{i=1}^{n} \exp\left(-\frac{|f_i - u_i|^2}{2\sigma^2}\right) \prod_{i=1}^{n-1} \exp\left(- \lambda|u_i - u_{i+1}|\right)$$

Rather than maximizing this probability distribution, one can equivalently minimize its negative logarithm (because the logarithm is strictly monotonous).

It is given by the energy

$$E(u) = -\log \mathcal{P}(u|f) = \sum_{i=1}^{n} \frac{|f_i - u_i|^2}{2\sigma^2} + \lambda \sum_{i=1}^{n-1} |u_i - u_{i+1}| + \text{const.}$$

MAP Estimation in the Continuous Setting

Similarly one can define Bayesian MAP optimization in the continuous setting, where the likelihood is given by:

$$\mathcal{P}(f|u) \propto \exp\left(-\int \frac{|f(x)-u(x)|^2}{2\sigma^2}dx\right),$$

and the prior is given by

$$\mathcal{P}(u) \propto \exp\left(-\lambda \int |\nabla u(x)| dx\right).$$

Thus the data term in variational methods corresponds to the likelihood, whereas the regularizer corresponds to the prior:

$$E(u) = -\log \mathcal{P}(u|f) = \int \frac{|f(x)-u(x)|^2}{2\sigma^2}dx + \lambda \int |\nabla u(x)| dx + \text{const.}$$

A systematic derivation of probability distributions on infinite-dimensional spaces requires a more formal derivation (introduction of measures etc). This is beyond our scope.

Conclusion on MAP Estimation

By invoking a Bayesian MAP rationale, one can formulate lots of computer vision problem under the form

$$\min_u D(u, f) + \lambda R(u)$$

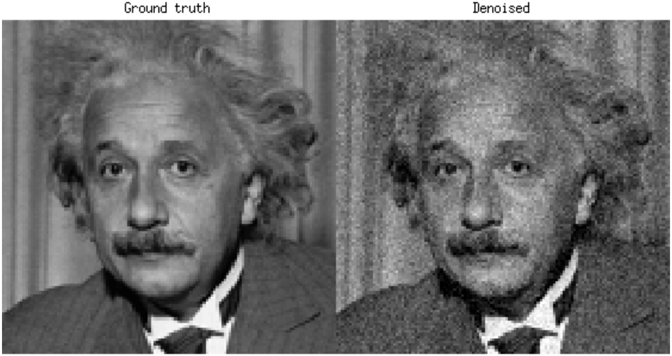
with an “automated” method for selecting D, R and λ.

A Few Classic Fidelity Terms

Gaussian noise is fine for modeling “small” perturbations:

$$\mathcal{P}(f|u) \propto \exp\left(-\int \frac{|f(x)-u(x)|^2}{2\sigma^2} dx\right)$$

$$\Rightarrow \min_u \underbrace{\int |f(x)-u(x)|^2 \, \mathrm{d}x}_{:=\|f-u\|_2^2} + \lambda R(u)$$

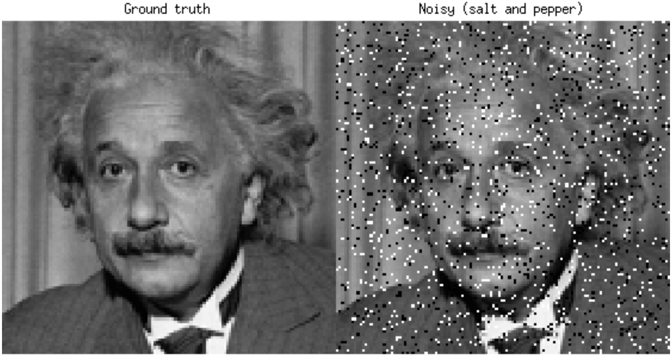


A Few Classic Fidelity Terms

Laplace noise is fine for modeling "impulsive" perturbations:

$$\mathcal{P}(f|u) \propto \exp\left(-\int \frac{|f(x)-u(x)|}{\sigma} dx\right)$$

$$\Rightarrow \underbrace{\min_u \int |f(x)-u(x)|\,\mathrm{d}x}_{:=\|f-u\|_1} + \lambda R(u)$$

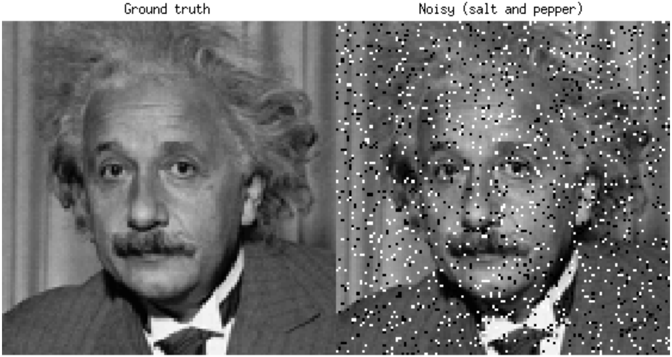


A Few Classic Fidelity Terms

Cauchy noise is even better for modeling “impulsive” perturbations:

$$\mathcal{P}(f|u) \propto \int \frac{1}{|f(x)-u(x)|^2+\sigma^2} dx$$

$$\Rightarrow \min_u \int \log\left(|f(x)-u(x)|^2+\sigma^2\right) \,\mathrm{d}x + \lambda R(u)$$

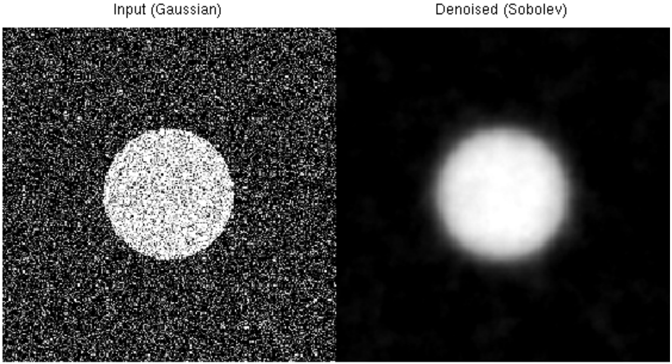


A Few Classic Regularizers

Sobolev regularization tends to favor smoothness:

$$\mathcal{P}(u) \propto \exp\left(-\lambda \int |\nabla u(x)|^2 dx\right)$$

$$\Rightarrow \min_u D(u,f) + \int \|\nabla u(x)\|^2 \,\mathrm{d}x$$

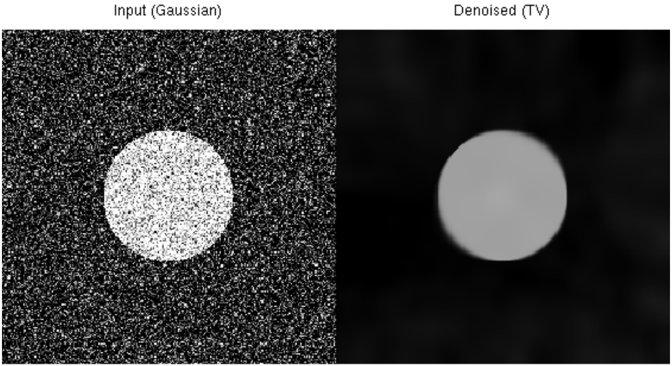


A Few Classic Regularizers

Total variation (TV) tends to favor piecewise constantness:

$$\mathcal{P}(u) \propto \exp\left(-\lambda \int |\nabla u(x)| dx\right)$$

$$\Rightarrow \min_u D(u, f) + \int \|\nabla u(x)\| \, \mathrm{d}x$$



Why TV Preserves Edges and Sobolev does not

Consider the following step function:

$$u(x) = \begin{cases} 0 \text{ if } x \leqslant h \\ \frac{a}{2h}x + \frac{a}{2} \text{ if } x \in [-h, h] \\ a \text{ if } x \geqslant h \end{cases}$$

Now remark that

$$\int_{\mathbb{R}} |u'(x)|^p dx = (2h)^{1-p} a^p$$

tends to infinity for $h \to 0$ if $p > 1$!!! Means: Sobolev regularization will never ever allow an edge.

But no problem if $p = 1$ (TV) !

Overview

Image Restoration: Motion Blur

Assume the camera lens opens instantly and remains open during the time interval $[0, T]$ in which the camera moves with constant velocity V in x-direction. The observed brightness is

$$g(x, y) = \frac{1}{T} \int_0^T f(x - Vt, y) dt.$$

Inserting $x' \equiv Vt$, we get a convolution

$$g(x, y) = \frac{1}{VT} \int_0^{VT} f(x - x', y) dx' = \int_{-\infty}^{\infty} f(x - x', y - y') h(x', y') dx' dy',$$

with the anisotropic blur kernel:

$$h(x', y') = \frac{1}{VT} \cdot \delta(y') \cdot \chi_{[0, VT]}(x'),$$

and

$$\chi_{[a,b]}(x') = \begin{cases} 1, & \text{if } x' \in [a, b] \\ 0, & \text{else} \end{cases} \qquad \text{(box filter)}$$

Example: Motion Blur

Original

Motion-blurred

(Author: D. Cremers)

Image Restoration: Defocus Blur

Defocus blur arises with real (in particular thick) lenses because structures are increasingly blurred, the further they are from the focal plane.

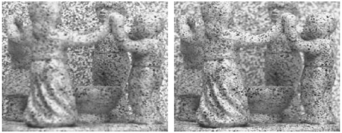
Depending on the focal setting and the depth of the scene, we will therefore observe a space-varying blur which allows us to infer the local depth (shape from focus / defocus).

Scene captured with three different focal settings.

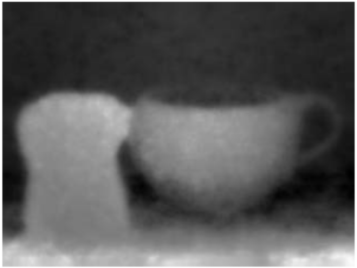
(Source: Favaro, Soatto, PAMI 2005)

Image Restoration: Defocus Blur

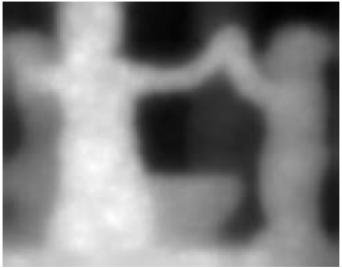
images with different focus



images with different focus



depth reconstruction



depth reconstruction

(Favaro et al., IEEE T. on PAMI 2008)

Overview

Image Restoration: Super Resolution

Super resolution from video exploits the redundancy available in multiple images. We assume that each image f_i is a blurred and downsampled version of a high-resolution scene.

We can try to recover a high-resolution image u with a variational approach of the form:

$$\min_u \sum_{i=1}^{n} \int |A(u \circ w_i) - f_i| \, dx \; + \; \lambda \int |\nabla u| \, dx.$$

The deformation field $w_i : \Omega \to \Omega$ models the warping from the original scene into image i, and A is a linear operator modeling the blurring and downsampling. Again, the variational approach aims at inverting an image formation process of the form:

$$f_i = A(u \circ w_i) + \eta,$$

which states that the observed image is obtained from the "true" image by nonrigid deformation, blurring and downsampling plus additive Poisson-distributed noise η.

Image Restoration: Super Resolution

One of several input images

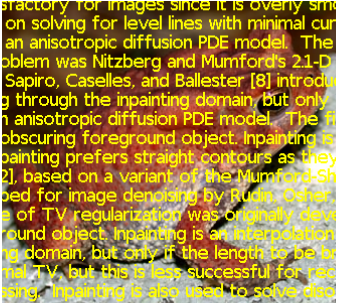
Superresolution estimate

(Schoenemann, Cremers, IEEE T. on Image Processing 2012)

Overview

Image Restoration: Inpainting

Image inpainting is a particular image restoration technique which explicitly handles (interpolate and / or extrapolate) missing data.



Corrupted

Denoised

Assume $f : \Omega \subset \mathbb{R}^2 \to \mathbb{R}$ a graylevel image, but only $\Omega_D \subset \Omega$ is "reliable". Then, denoising (or deblurring, etc.) should not use the f-data over $\Omega \backslash \Omega_D$. The standard TV-inpainting model is then:

$$\min_u \int_{\Omega_D} |u - f|^2 \, dx \; + \; \lambda \int_{\Omega} |\nabla u| \, dx.$$

Limits of TV-inpainting

TV-inpainting is a very naive interpolation technique which does not transport texture...

Variations around the Inpainting Model

- Compression by Diffusion

- Poisson Image Editing [Perez et al, SIGGRAPH 2003]

How would you do that ?

Overview

Gradient Descent for the L2-TV (ROF) Model

Consider the generic L2-TV (ROF) restoration problem:

$$\min_{u\colon \Omega\subset\mathbb{R}^2\to\mathbb{R}} \int_\Omega \frac{1}{2}\left((Ku)(x) - f(x)\right)^2 + \lambda|\nabla u(x)|\,\mathrm{d}x$$

(denoising: $K = \mathrm{id}$; deblurring: K = Gaussian kernel, super-resolution: K = zoom kernel, etc.)

Its first-order optimality condition is the Euler-Lagrange equation

$$K^\top(Ku - f) - \lambda\nabla\cdot\left(\frac{\nabla u}{|\nabla u|}\right) \quad \text{over } \Omega,$$

with Neumann or Dirichlet boundary conditions on $\partial\Omega$.

Starting from $u(x,0) = f(x)\ \forall x \in \Omega$, optimization can be carried out by gradient descent:

$$\partial_t u(x,t) = -K^\top\left(Ku(x,t) - f(x)\right) + \lambda\nabla_x\cdot\left(\frac{\nabla_x u(x,t)}{|\nabla_x u(x,t)|}\right),$$
$$\forall (x,t) \in \Omega\times[0,+\infty).$$

Main Issue in Gradient Descent for the L2-TV (ROF) Model

When implementing the gradient descent

$$\partial_t u(x,t) = -K^\top \left(Ku(x,t) - f(x)\right) + \lambda \nabla_x \cdot \left(\frac{\nabla_x u(x,t)}{|\nabla_x u(x,t)|} \right),$$
$$\forall (x,t) \in \Omega \times [0,+\infty),$$

one must be careful to avoid division by zero which occurs due to the factor $|\nabla_x u(x,t)|$ (infinite diffusivity if there is no edge).

In practice, we need to smooth a bit this term:

$$\frac{1}{|\nabla_x u(x,t)|} \approx \frac{1}{|\nabla_x u(x,t)|_\epsilon} := \frac{1}{\sqrt{|\nabla_x u(x,t)|^2 + \epsilon}}$$

or

$$\frac{1}{|\nabla_x u(x,t)|} \approx \frac{1}{|\nabla_x u(x,t)|_\mu} := \frac{1}{\max\{\mu, |\nabla_x u(x,t)|\}}$$

with $\epsilon, \mu > 0$, small (e.g. 10^{-3})

The Choice of the Smoothing Matters

Example with $\epsilon = \mu = 10^{-3}$

The Choice of the Smoothing Matters

Example with $\epsilon = \mu = 10^{-2}$

The Choice of the Smoothing Matters

Example with $\epsilon = \mu = 10^{-1}$

What we are Actually Doing...

The first “smoothed” gradient descent

$$\partial_t u(x,t) = -K^\top \left(Ku(x,t) - f(x)\right) + \lambda \nabla_x \cdot \left(\frac{\nabla_x u(x,t)}{|\nabla_x u(x,t)|_\epsilon} \right), \quad \forall (x,t) \in \Omega \times [0,+\infty),$$

is exactly the gradient descent for the “smoothed” functional

$$\min_{u\colon \Omega\subset\mathbb{R}^2\to\mathbb{R}} \int_\Omega \frac{1}{2} \left((Ku)(x) - f(x)\right)^2 + \lambda |\nabla_x u(x)|_\epsilon \, \mathrm{d}x.$$

What we are Actually Doing...

The second "smoothed" gradient descent

$$\partial_t u(x,t) = -K^\top \left(Ku(x,t) - f(x)\right) + \lambda \nabla_x \cdot \left(\frac{\nabla_x u(x,t)}{|\nabla_x u(x,t)|_\mu} \right), \quad \forall (x,t) \in \Omega \times [0,+\infty),$$

is exactly the gradient descent for the "smoothed" functional

$$\min_{u\colon \Omega\subset\mathbb{R}^2\to\mathbb{R}} \int_\Omega \frac{1}{2}\left((Ku)(x) - f(x)\right)^2 + \lambda \frac{|\nabla u(x)|^2}{|\nabla_x u(x)|_\mu} \, \mathrm{d}x.$$

What we are Actually Doing...

The Moreau-Yosida regularization of a function $f(x)$ is defined as follows:

$$M_\mu f(x) = \inf_y \left\{ f(y) + \frac{1}{2\mu} \|x - y\|^2 \right\}$$

Under mild conditions, f and its Moreau envelope have the same minimizers. But minimizing the Moreau envelope is often much easier (it is smooth, even if f is not). $\rightarrow$ we often rather tackle the Moreau envelope if the optimization problem is nonsmooth. This is the basic idea of proximal optimization.

For the absolute value $f(x) = |x|$, we get (up to constants) $M_\mu f(x) = \frac{x^2}{|x|_\mu}$ (i.e., the Huber loss), hence minimizing the smoothed functional

$$\min_{u:\,\Omega\subset\mathbb{R}^2\rightarrow\mathbb{R}} \int_\Omega \frac{1}{2}\left((Ku)(x) - f(x)\right)^2 + \lambda \frac{|\nabla u(x)|^2}{|\nabla_x u(x)|_\mu} \,\mathrm{d}x$$

is very closely related to the original (non-smoothed) variational problem (not just a computer scientist hack...).

Explicit Time Gradient Descent

We can now discretize the gradient descent equation

$$\partial_t u(x,t) = -K^\top \left(Ku(x,t) - f(x)\right) + \lambda \nabla_x \cdot \left(\frac{\nabla_x u(x,t)}{|\nabla_x u(x,t)|_\mu} \right), \quad \forall (x,t) \in \Omega \times [0,+\infty),$$

wrt time t using forward finite differences i.e.,

$$\partial_t u(x,t) = \frac{u(x,t+1) - u(x,t)}{\delta_t},$$

with some fixed stepsize $\delta_t > 0$.

This yields the following algorithm:

$$u^{(0)} = f$$

$$u^{(t+1)} = u^{(t)} - \delta_t \left(K^\top \left(Ku^{(t)} - f \right) - \lambda \nabla \cdot \frac{\nabla u^{(t)}}{|\nabla u^{(t)}|_\mu} \right), \quad t \in \{1, 2, \dots\}$$

This works, but descent has to be slow (low δ_t)

Lagged Diffusivity (Implicit Time Gradient Descent)

To make things more stable, we usually prefer to freeze only the diffusivity during descent, i.e.:

$$u^{(0)} = f$$
$$u^{(t+1)} = u^{(t)} - \delta_t \left(K^\top \left(K u^{(t+1)} - f \right) - \lambda \nabla \cdot \frac{\nabla u^{(t+1)}}{|\nabla u^{(t)}|_\mu} \right), \quad t \in \{1, 2, \ldots\},$$

which requires a linear system to be solved at each update:

$$\left(\mathrm{id} + \delta_t K^\top K - \delta_t \lambda \nabla \cdot \left(\frac{1}{|\nabla u^{(t)}|_\mu} \nabla \right) \right) u^{(t+1)} = u^{(t)} + \delta_t K^\top f$$

Typically much larger stepsizes are allowed, which makes things way faster and removes the need for tedious tuning (or linesearch).

The Choice of the Stepsize Matters

Example with $\delta_t = 0.02$

The Choice of the Stepsize Matters

Example with $\delta_t = 0.2$

The Choice of the Stepsize Matters

Example with $\delta_t = 2$

Gradient Descent Process for the Inpainting + Denoising Task