Total variation regularization by iteratively reweighted least squares on Hadamard spaces and Riemannian manifolds

Markus Sprecher, ETH Zürich
Joint work with Prof. Dr. Philipp Grohs, ETH Zürich

August 20, 2015

Problem description

We want to denoise and/or inpaint Images $\{1, \ldots, m\} \times \{1, \ldots, n\} \to X$, e. g.

Figure: Grayscale and RGB-image, $X = \mathbb{R}$ resp. $\mathbb{R}^3$.

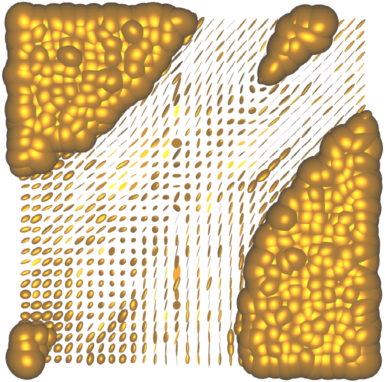


Figure: diffusion tensor magnetic resonance imaging (DT-MRI), $X = SPD(3)$, the space of 3×3 symmetric positive definite matrices.

Outline

Outline

1. Case $X = \mathbb{R}$
2. General Case

Outline

Outline

1. Case $X = \mathbb{R}$
2. General Case
3. Experiments
4. Comparison to proximal point method

Total variation regularization for real-valued functions

Consider a noisy image $u^0 \colon \{1, \ldots, m\} \times \{1, \ldots, n\} \to \mathbb{R}$. We can construct a restoration by minimizing the functional

$$J(u) := \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{n} (u_{i,j} - u^0_{i,j})^2 + \lambda TV(u)$$

where $\lambda > 0$ and

$$TV(u) := \sum_{i=1}^{m-1} \sum_{j=1}^{n} |u_{i+1,j} - u_{i,j}| + \sum_{i=1}^{m} \sum_{j=1}^{n-1} |u_{i,j+1} - u_{i,j}|.$$

This functional was introduced by Rudin, Osher and Fatemi in 1992 [3].

Properties of J:

Properties of J:

- Convex $\Rightarrow$ unique minimizer.
- Not differentiable (due to modulus)$\Rightarrow$ can not apply Newton

Properties of J:

- Convex $\Rightarrow$ unique minimizer.
- Not differentiable (due to modulus)$\Rightarrow$ can not apply Newton

Idea: Replace $|u_{i+1,j} - u_{i,j}|$ by $w_{i,j}(u_{i+1,j} - u_{i,j})^2$ where $w_{i,j} \approx |u_{i+1,j} - u_{i,j}|^{-1}$.

Properties of J:

- Convex $\Rightarrow$ unique minimizer.
- Not differentiable (due to modulus)$\Rightarrow$ can not apply Newton

Idea: Replace $|u_{i+1,j} - u_{i,j}|$ by $w_{i,j}(u_{i+1,j} - u_{i,j})^2$ where $w_{i,j} \approx |u_{i+1,j} - u_{i,j}|^{-1}$.
Choose $u^{(0)}$ and define an iteration by

$$\begin{aligned} w_{i,j}^{(k+1)} &:= \left((u_{i+1,j}^{(k)} - u_{i,j}^{(k)})^2 + \epsilon^2\right)^{-\frac{1}{2}} \text{ and} \\ u^{(k+1)} &:= \operatorname*{argmin}_{u \in \mathbb{R}^{m\times n}} \sum_{i=1}^{m}\sum_{j=1}^{n}(u_{i,j} - u_{i,j}^{0})^2 \\ &\quad +\lambda \sum_{i=1}^{m-1}\sum_{j=1}^{n} w_{i,j}^{(k+1)}(u_{i+1,j} - u_{i,j})^2 + \dots \end{aligned}$$

In each step a quadratic functional has to be optimized. To do that we only need to solve a sparse linear system.

In each step a quadratic functional has to be optimized. To do that we only need to solve a sparse linear system.

Sketch of analysis of IRLS:

In each step a quadratic functional has to be optimized. To do that we only need to solve a sparse linear system.

Sketch of analysis of IRLS:

One proves that the iteration converges to the unique minimizer of

$$J^{\epsilon}(u) := \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n}(u_{i,j}-u_{i,j}^{0})^2+\lambda\sum_{i=1}^{m-1}\sum_{j=1}^{n}\sqrt{(u_{i+1,j}-u_{i,j})^2+\epsilon^2}\ldots$$

In each step a quadratic functional has to be optimized. To do that we only need to solve a sparse linear system.

Sketch of analysis of IRLS:

One proves that the iteration converges to the unique minimizer of

$$J^{\epsilon}(u) := \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n}(u_{i,j}-u^{0}_{i,j})^2+\lambda\sum_{i=1}^{m-1}\sum_{j=1}^{n}\sqrt{(u_{i+1,j}-u_{i,j})^2+\epsilon^2}\ldots$$

The IRLS algorithm can be reinterpreted as alternating minimization of

$$\begin{aligned}\tilde{J}^{\epsilon}(w,u) \quad := \quad & \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n}(u_{i,j}-u^{0}_{i,j})^2 \\ & +\frac{1}{2}\lambda\sum_{i=1}^{m-1}\sum_{j=1}^{n} w_{i,j}((u_{i+1,j}-u_{i,j})^2+\epsilon^2)+(w_{i,j})^{-1}\ldots\end{aligned}$$

Connections to J^ϵ:

- $J^\epsilon(u) = \tilde{J}^\epsilon(w^\epsilon(u), u)$. Hence $J^\epsilon(u^{(k)})$ is nonincreasing.
- (w, u) is a critical point of $\tilde{J}^\epsilon$ iff u is a critical point of J^ϵ and $w = w^\epsilon(u)$. Hence $\tilde{J}^\epsilon$ has a unique critical point.

Connections to J^ϵ:

- $J^\epsilon(u) = \tilde{J}^\epsilon(w^\epsilon(u), u)$. Hence $J^\epsilon(u^{(k)})$ is nonincreasing.
- (w, u) is a critical point of $\tilde{J}^\epsilon$ iff u is a critical point of J^ϵ and $w = w^\epsilon(u)$. Hence $\tilde{J}^\epsilon$ has a unique critical point.

Theorem

The sequence $(u^{(k)})_{k\in\mathbb{N}}$ generated by IRLS converges to the unique minimizer of J^ϵ.

Sketch of proof:

- $J^\epsilon(u^{(k)})$ convergences

Connections to J^ϵ:

- $J^\epsilon(u) = \tilde{J}^\epsilon(w^\epsilon(u), u)$. Hence $J^\epsilon(u^{(k)})$ is nonincreasing.
- (w, u) is a critical point of $\tilde{J}^\epsilon$ iff u is a critical point of J^ϵ and $w = w^\epsilon(u)$. Hence $\tilde{J}^\epsilon$ has a unique critical point.

Theorem

The sequence $(u^{(k)})_{k\in\mathbb{N}}$ generated by IRLS converges to the unique minimizer of J^ϵ.

Sketch of proof:

- $J^\epsilon(u^{(k)})$ convergences
- $(w^{(k)}, u^{(k)})$ bounded $\Rightarrow \exists$ convergent subsequence

Connections to J^ϵ:

- $J^\epsilon(u) = \tilde{J}^\epsilon(w^\epsilon(u), u)$. Hence $J^\epsilon(u^{(k)})$ is nonincreasing.
- (w, u) is a critical point of $\tilde{J}^\epsilon$ iff u is a critical point of J^ϵ and $w = w^\epsilon(u)$. Hence $\tilde{J}^\epsilon$ has a unique critical point.

Theorem

The sequence $(u^{(k)})_{k\in\mathbb{N}}$ generated by IRLS converges to the unique minimizer of J^ϵ.

Sketch of proof:

- $J^\epsilon(u^{(k)})$ convergences
- $(w^{(k)}, u^{(k)})$ bounded $\Rightarrow \exists$ convergent subsequence
- Show that the limit of this convergent subsequence is a critical point of $\tilde{J}^\epsilon$.

Connections to J^ϵ:

- $J^\epsilon(u) = \tilde{J}^\epsilon(w^\epsilon(u), u)$. Hence $J^\epsilon(u^{(k)})$ is nonincreasing.
- (w, u) is a critical point of $\tilde{J}^\epsilon$ iff u is a critical point of J^ϵ and $w = w^\epsilon(u)$. Hence $\tilde{J}^\epsilon$ has a unique critical point.

Theorem

The sequence $(u^{(k)})_{k\in\mathbb{N}}$ generated by IRLS converges to the unique minimizer of J^ϵ.

Sketch of proof:

- $J^\epsilon(u^{(k)})$ convergences
- $(w^{(k)}, u^{(k)})$ bounded $\Rightarrow \exists$ convergent subsequence
- Show that the limit of this convergent subsequence is a critical point of $\tilde{J}^\epsilon$.
- Show that $(u^{(k)})$ convergence to the critical point of J^ϵ.

What if $X \neq \mathbb{R}$?

What if $X \neq \mathbb{R}$?
If (X, d) is a metric space we can replace $|u - v|$ by $d(u, v)$, i.e.

$$J(u) := \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{n} d^2(u_{i,j}, u_{i,j}^0) + \lambda TV(u)$$

where $\lambda > 0$ and

$$TV(u) := \sum_{i=1}^{m-1} \sum_{j=1}^{n} d(u_{i+1,j}, u_{i,j}) + \sum_{i=1}^{m} \sum_{j=1}^{n-1} d(u_{i,j+1}, u_{i,j}).$$

What if $X \neq \mathbb{R}$?
If (X, d) is a metric space we can replace $|u - v|$ by $d(u, v)$, i.e.

$$J(u) := \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{n} d^2(u_{i,j}, u^0_{i,j}) + \lambda TV(u)$$

where $\lambda > 0$ and

$$TV(u) := \sum_{i=1}^{m-1} \sum_{j=1}^{n} d(u_{i+1,j}, u_{i,j}) + \sum_{i=1}^{m} \sum_{j=1}^{n-1} d(u_{i,j+1}, u_{i,j}).$$

Is J convex?

What if $X \neq \mathbb{R}$?
If (X, d) is a metric space we can replace $|u - v|$ by $d(u, v)$, i.e.

$$J(u) := \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{n} d^2(u_{i,j}, u_{i,j}^0) + \lambda TV(u)$$

where $\lambda > 0$ and

$$TV(u) := \sum_{i=1}^{m-1} \sum_{j=1}^{n} d(u_{i+1,j}, u_{i,j}) + \sum_{i=1}^{m} \sum_{j=1}^{n-1} d(u_{i,j+1}, u_{i,j}).$$

Is J convex?
If X is a geodesic space and d is convex, (i.e. if $t \mapsto d(\gamma_1(t), \gamma_2(t))$ is convex for geodesics $\gamma_1, \gamma_2 \colon [0, 1] \to X$) then J is also convex.

What if $X \neq \mathbb{R}$?
If (X, d) is a metric space we can replace $|u - v|$ by $d(u, v)$, i.e.

$$J(u) := \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{n} d^2(u_{i,j}, u_{i,j}^0) + \lambda TV(u)$$

where $\lambda > 0$ and

$$TV(u) := \sum_{i=1}^{m-1} \sum_{j=1}^{n} d(u_{i+1,j}, u_{i,j}) + \sum_{i=1}^{m} \sum_{j=1}^{n-1} d(u_{i,j+1}, u_{i,j}).$$

Is J convex?
If X is a geodesic space and d is convex, (i.e. if $t \mapsto d(\gamma_1(t), \gamma_2(t))$ is convex for geodesics $\gamma_1, \gamma_2 \colon [0, 1] \to X$) then J is also convex.
However there are spaces where d is **not** convex, e.g. the sphere.

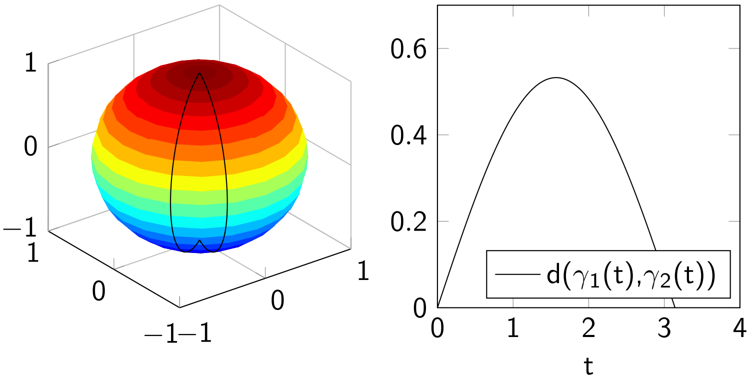


Figure: Distance function on the sphere is not convex

Hadamard spaces

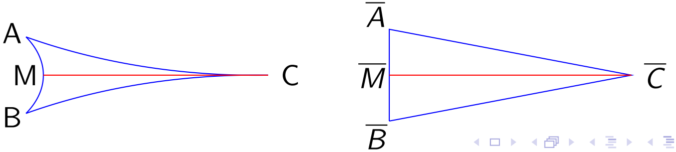
However, for Hadamard spaces, a.k.a. spaces of nonpositive curvature the distance function is convex.

Definition

A complete metric space (X, d) is called a Hadamard space if for all $A, B \in X$ there exist $M \in X$ such that for all $C \in X$ we have

$$d^2(C, M) \leqslant \frac{1}{2} d^2(C, A) + \frac{1}{2} d^2(C, B) - \frac{1}{4} d^2(A, B). \tag{1}$$

RHS in (1) is length of the median of the comparison triangle in $\mathbb{R}^2$.



The optimization problem for an arbitrary metric space X is

$$
\begin{aligned}
w_{i,j}^{(k+1)} &:= \left(d^2(u_{i+1,j}^{(k)}, u_{i,j}^{(k)}) + \epsilon^2\right)^{-\frac{1}{2}} \\
u^{(k+1)} &:= \operatorname*{argmin}_{u \in X^{m\times n}} \sum_{i=1}^{m} \sum_{j=1}^{n} d^2(u_{i,j}, u_{i,j}^0) \\
&\quad + \lambda \sum_{i=1}^{m-1} \sum_{j=1}^{n} w_{i,j}^{(k+1)} d^2(u_{i+1,j}, u_{i,j}) + \ldots .
\end{aligned}
$$

The optimization problem for an arbitrary metric space X is

$$\begin{aligned} w_{i,j}^{(k+1)} &:= \left(d^2(u_{i+1,j}^{(k)}, u_{i,j}^{(k)}) + \epsilon^2\right)^{-\frac{1}{2}} \\ u^{(k+1)} &:= \operatorname*{argmin}_{u \in X^{m \times n}} \sum_{i=1}^{m} \sum_{j=1}^{n} d^2(u_{i,j}, u_{i,j}^0) \\ &\quad + \lambda \sum_{i=1}^{m-1} \sum_{j=1}^{n} w_{i,j}^{(k+1)} d^2(u_{i+1,j}, u_{i,j}) + \dots . \end{aligned}$$

Hence in every step we need to minimize a functional of the form

$$(u_l)_{l=1}^N \mapsto \sum_{l=1}^{N} d^2(u_l, v_l) + \sum_{l=1}^{N} \sum_{l'=1}^{N} W_{l,l'} d^2(u_l, u_{l'}).$$

where $(v_l)_{l=1}^N \subset X$ and $W_{l,l'} \geqslant 0$.

Riemannian exponential map

A basic problem when trying to generalize concepts to manifolds is that we can not 'add' two elements of the manifold. However we can add to an element $x \in M$ a tangent vector $r \in T_x M$ by the Riemannian exponential map.

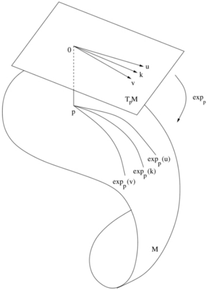


Figure: Exponential map

Riemannian Newton method

To solve an optimization problem $\operatorname{argmin}_{u \in M} f(u)$ where $f : M \to \mathbb{R}$ one can use the Riemannian Newton method. The iteration is

$$\phi(x) := \exp_x \left(-\left(\mathrm{Hess} f(x)\right)^{-1} [\mathrm{grad} f(x)] \right),$$

where grad and Hess are the intrinsic gradient and Hessian, i.e. the unique functions such that

$$f(\exp_x(r)) = f(x) + \langle \mathrm{grad} f(x), r \rangle + \frac{1}{2} \langle r, \mathrm{Hess} f(x)[r] \rangle + \mathcal{O}(|r|^3),$$

where $r \in T_x M$.

Riemannian Newton method

To solve an optimization problem $\operatorname{argmin}_{u \in M} f(u)$ where $f: M \to \mathbb{R}$ one can use the Riemannian Newton method. The iteration is

$$\phi(x) := \exp_x \left(- \left(\mathrm{Hess} f(x)\right)^{-1} [\mathrm{grad} f(x)] \right),$$

where grad and Hess are the intrinsic gradient and Hessian, i.e. the unique functions such that

$$f(\exp_x(r)) = f(x) + \langle \mathrm{grad} f(x), r \rangle + \frac{1}{2} \langle r, \mathrm{Hess} f(x)[r] \rangle + \mathcal{O}(|r|^3),$$

where $r \in T_x M$.
The Riemannian Newton method has quadratic convergence (Absil [1]).

Back to the sphere

Why is the sphere case important?

Back to the sphere

Why is the sphere case important?
An RGB image $(i,j) \mapsto u_{i,j} \in \mathbb{R}^3$ can be separated into its brightness $(i,j) \mapsto |u_{i,j}| \in \mathbb{R}$ and color part $(i,j) \mapsto u_{i,j}/|u_{i,j}| \in S^2$.

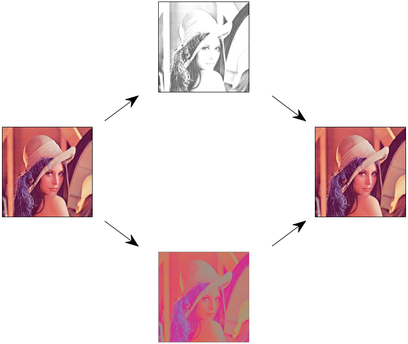


Figure: Separation of RGB image into brightness and color part.

As we have seen our functional is not convex when X is the sphere. However in our example we can restrict to a half sphere.

As we have seen our functional is not convex when X is the sphere. However in our example we can restrict to a half sphere.

Lemma

If X is an open half sphere, d the spherical distance and $(v_l)_{l=1}^N \subset X$ the functional

$$(u_l)_{l=1}^N \mapsto \sum_{l=1}^N d^2(u_l, v_l) + \sum_{l=1}^N \sum_{l'=1}^N W_{l,l'} d^2(u_l, u_{l'}).$$

is locally convex at every critical point in X^N.

Therefore any critical point is a local minimizer.

As we have seen our functional is not convex when X is the sphere. However in our example we can restrict to a half sphere.

Lemma

If X is an open half sphere, d the spherical distance and $(v_l)_{l=1}^N \subset X$ the functional

$$(u_l)_{l=1}^N \mapsto \sum_{l=1}^N d^2(u_l, v_l) + \sum_{l=1}^N \sum_{l'=1}^N W_{l,l'} d^2(u_l, u_{l'}).$$

is locally convex at every critical point in X^N.

Therefore any critical point is a local minimizer.
This can be used to prove that there is only one minimizer

In $1d$ between any two local minimal point of a continuous function there is a local maximum point.

In $1d$ between any two local minimal point of a continuous function there is a local maximum point.

Theorem (Poincaré-Hopf [2])

Let M be a compact manifold with boundary and $U\colon M \to TM$ a vector field on M such that U is pointing outward on the boundary of M. Assume that U has a continuous derivative DU, all zeros of U are isolated and $DU(z)$ is invertible for all zeros $z \in M$ of U. Then U has finitely many zeros $z_1, \ldots, z_n \in M$ and

$$\sum_{i=1}^{n} sign(\det(DU(z_i))) = \chi(M),$$

where $\chi(M)$ is the Euler characteristic of M.

Colorization

We assume that we know the brightness but the color part of every pixel is only known with probability 1%.

- Detect edges using Canny edge detector on the brightness image.
- Construct a first guess using scattered interpolation.
- Minimize a weighted TV-functional (10^{-2} at edges and 1 everywhere else)

Figure: Colorization.

SPD denoising: Riemannian vs. Euclidean

The Riemannian metric on $SPD(n)$ is usually defined by $\langle X, Y \rangle_A := \mathrm{trace}(A^{-1}XA^{-1}Y)$ which induces the metric $d(A, B) = \|\log(A^{-\frac{1}{2}}BA^{-\frac{1}{2}})\|_F$. Why not just regard $SPD(n)$ as a subspace of the Euclidean space $\mathbb{R}^{n^2}$ and use the standard metric $d_{euc}(A, B) = \|A - B\|_F$?

SPD denoising: Riemannian vs. Euclidean

The Riemannian metric on $SPD(n)$ is usually defined by $\langle X, Y\rangle_A := \mathrm{trace}(A^{-1}XA^{-1}Y)$ which induces the metric $d(A, B) = \|\log(A^{-\frac{1}{2}}BA^{-\frac{1}{2}})\|_F$. Why not just regard $SPD(n)$ as a subspace of the Euclidean space $\mathbb{R}^{n^2}$ and use the standard metric $d_{euc}(A, B) = \|A - B\|_F$?

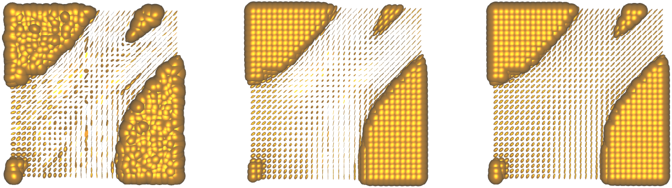
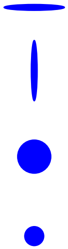


Figure: From left to right: Noisy image, recovered image with SPD norm, recovered image with Euclidean norm.

TV-denoising is similar to averaging the data.

TV-denoising is similar to averaging the data.

$$\begin{aligned}
A &= \begin{pmatrix} 1 & 0 \\ 0 & \epsilon \end{pmatrix} & \det(A) &= \epsilon \\
B &= \begin{pmatrix} \epsilon & 0 \\ 0 & 1 \end{pmatrix} & \det(B) &= \epsilon \\
av_{Euc}(A,B) &= \begin{pmatrix} \frac{1}{2} + \frac{\epsilon}{2} & 0 \\ 0 & \frac{1}{2} + \frac{\epsilon}{2} \end{pmatrix} & \det(av_{Euc}(A,B)) &\approx \frac{1}{4} \\
av_{SPD}(A,B) &= \begin{pmatrix} \sqrt{\epsilon} & 0 \\ 0 & \sqrt{\epsilon} \end{pmatrix} & \det(av_{SPD}(A,B)) &= \epsilon
\end{aligned}$$



Sketch of PPM (Weinmann et al. [4])

Idea: Solve the gradient flow $\dot{u} = -\nabla J(u)$. $u_\infty := \lim_{t\to\infty} u(t)$ is the minimizer of J. Use implicit Euler and timesteps $(t_k)_{k\in\mathbb{N}}$ with $\sum_{k\in\mathbb{N}} t_k = \infty$ and $\sum_{k\in\mathbb{N}} t_k^2 < \infty$.

Sketch of PPM (Weinmann et al. [4])

Idea: Solve the gradient flow $\dot{u} = -\nabla J(u)$. $u_\infty := \lim_{t\to\infty} u(t)$ is the minimizer of J. Use implicit Euler and timesteps $(t_k)_{k\in\mathbb{N}}$ with $\sum_{k\in\mathbb{N}} t_k = \infty$ and $\sum_{k\in\mathbb{N}} t_k^2 < \infty$.
Problem: J not differentiable so

$$u^{(k+1)} = u^{(k)} + t_k \nabla J(u^{(k+1)})$$

does not make sense.

Sketch of PPM (Weinmann et al. [4])

Idea: Solve the gradient flow $\dot{u} = -\nabla J(u)$. $u_\infty := \lim_{t\to\infty} u(t)$ is the minimizer of J. Use implicit Euler and timesteps $(t_k)_{k\in\mathbb{N}}$ with $\sum_{k\in\mathbb{N}} t_k = \infty$ and $\sum_{k\in\mathbb{N}} t_k^2 < \infty$.
Problem: J not differentiable so

$$u^{(k+1)} = u^{(k)} + t_k \nabla J(u^{(k+1)})$$

does not make sense. However we can define

$$u^{(k+1)} = \mathrm{prox}_{t_k J}(u^{(k)}) := \operatorname*{argmin}_{u\in X^N} t_k J(u) + \frac{1}{2}\sum_{l=1}^{N} d^2(u_l, u_l^{(k)}). \quad (2)$$

By splitting J one can solve (2) exactly for each subproblem using Riemannian exponentials and its inverse.

Comparison of IRLS and PPM:

- In PPM we dont need a regularization parameter ϵ. However we need to choose a sequence of stepsizes $(t_k)_{k\in\mathbb{N}}$.
- IRLS is more complicated to implement as we need to compute second derivatives of the squared distance function and compute Hessian's.
- For IRLS we have linear convergence whereas for PPM we don't. The number of iteration needed is significantly less in IRLS.
- The major part of the computational time goes into computing the linear system of the Newton method for IRLS and into computing the Riemannian exponential and its inverse for PPM. The more expensive it is to compute the Riemannian exponential map and its inverse the faster IRLS is compared to PPM.
- IRLS can also deal with the isotropic TV

$$TV_{iso}(u) := \sum_{i=1}^{m}\sum_{j=1}^{n}\sqrt{d^2(u_{i+1,j},u_{i,j}) + d^2(u_{i,j+1},u_{i,j})}.$$

P. A. Absil, R. Mahony, and R. Sepulchre.
Optimization algorithms on matrix manifolds.
Princeton University Press, 2009.

J. Milnor.
Topology from the differentiable viewpoint.
Princeton University Press, 1976.

L. I. Rudin, S. Osher, and E. Fatemi.
Nonlinear total variation based noise removal algorithms.
Physica D, 60(1-4):259–268, 1992.

A. Weinmann, L. Demaret, and M. Storath.
Total variation regularization for manifold-valued data.
SIAM Journal on Imaging Sciences, 7(4):2226–2257, 2014.

Thank you for your attention!
Questions?