Medical Image Processing for Interventional Applications

Super-Resolution: Regularization

Online Course – Unit 23

Andreas Maier, Thomas Köhler, Frank Schebesch
Pattern Recognition Lab (CS 5)

Topics

Maximum A Posteriori Estimation

Image Priors

Summary

- Take Home Messages
- Further Readings

Super-Resolution and Ill-posed Problems

ML estimation on a synthetic example:

- Super-resolution is an **ill-posed** problem
  $\rightarrow$ ML reconstruction might lead to unstable solutions with amplified noise.

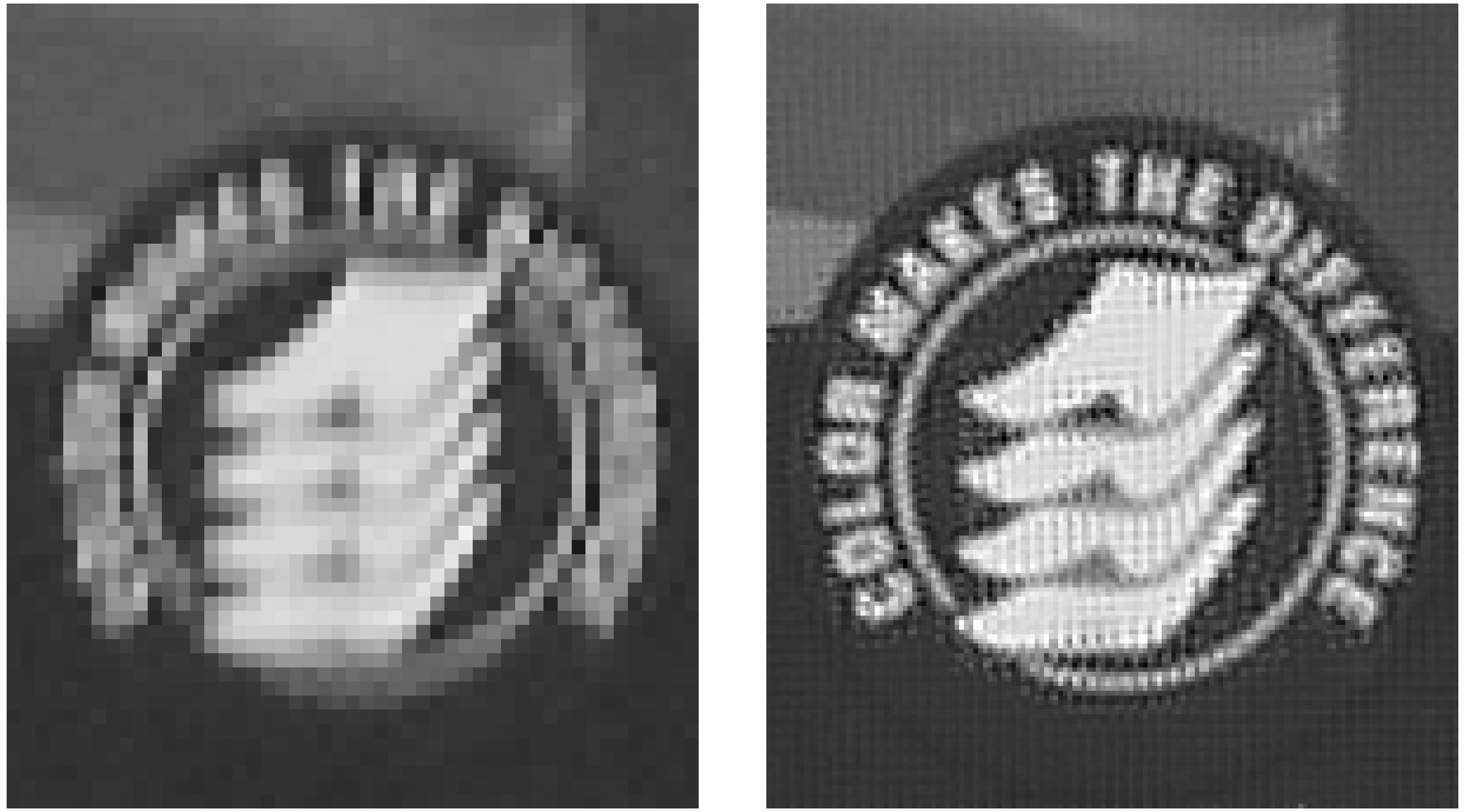

Figure 1: Example for insufficient result from ML estimation

Super-Resolution and Ill-posed Problems

ML estimation on a synthetic example:

- Super-resolution is an **ill-posed** problem $\rightarrow$ ML reconstruction might lead to unstable solutions with amplified noise.
- **Way out:** Incorporate prior knowledge into super-resolution algorithm.

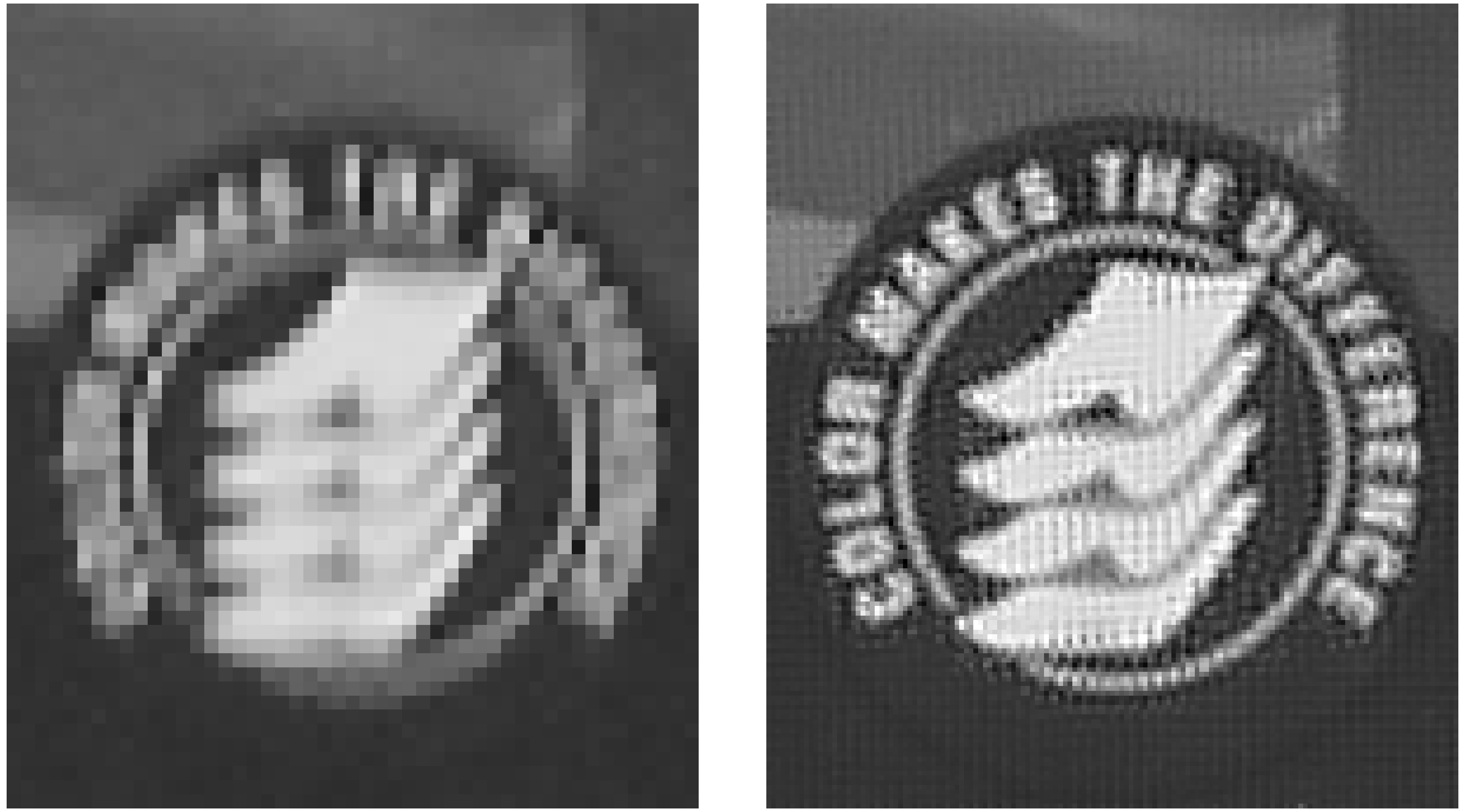

Figure 1: Example for insufficient result from ML estimation

Maximum A Posteriori Estimation

Extension of the ML estimation to MAP estimation:

- For maximum a posteriori (MAP) estimation, a suitable prior $p(\mathbf{x})$ is used.

Maximum A Posteriori Estimation

Extension of the ML estimation to MAP estimation:

- For maximum a posteriori (MAP) estimation, a suitable prior $p(\mathbf{x})$ is used.
- According to Bayes formula, we have:

$$\hat{\mathbf{x}}_{MAP} = \arg\max_{\mathbf{x}} p(\mathbf{x})p(\mathbf{y} \,|\, \mathbf{x}) = \arg\min_{\mathbf{x}} \{-\log p(\mathbf{x}) - \log p(\mathbf{y} \,|\, \mathbf{x})\} \,.$$

Maximum A Posteriori Estimation

Extension of the ML estimation to MAP estimation:

- For maximum a posteriori (MAP) estimation, a suitable prior $p(\mathbf{x})$ is used.
- According to Bayes formula, we have:

$$\hat{\mathbf{x}}_{MAP} = \arg\max_{\mathbf{x}} p(\mathbf{x})p(\mathbf{y}\,|\,\mathbf{x}) = \arg\min_{\mathbf{x}} \{-\log p(\mathbf{x}) - \log p(\mathbf{y}\,|\,\mathbf{x})\}\,.$$

- For additive Gaussian noise, MAP estimation is equivalent to:

$$\hat{\mathbf{x}}_{MAP} = \arg\min_{\mathbf{x}} \sum_{k=1}^{K} \left|\left|\mathbf{y}^{(k)} - \mathbf{W}^{(k)}\mathbf{x}\right|\right|_2^2 - \log p(\mathbf{x}) = \arg\min_{\mathbf{x}} ||\mathbf{y} - \mathbf{W}\mathbf{x}||_2^2 - \log p(\mathbf{x}).$$

Topics

Maximum A Posteriori Estimation

Image Priors

Summary

- Take Home Messages
- Further Readings

Image Prior and Regularization Term

Mathematical framework for image priors in super-resolution:

- We define the prior distribution according to the exponential form:

$$p(\boldsymbol{x}) = \frac{1}{Z} \exp\left(-\lambda R(\boldsymbol{x})\right),$$

where Z is a normalization constant, and $R(\boldsymbol{x})$ denotes a regularization term with regularization weight $\lambda \geq 0$.

Image Prior and Regularization Term

Mathematical framework for image priors in super-resolution:

- We define the prior distribution according to the exponential form:

$$p(\boldsymbol{x}) = \frac{1}{Z} \exp\left(-\lambda R(\boldsymbol{x})\right),$$

  where Z is a normalization constant, and $R(\boldsymbol{x})$ denotes a regularization term with regularization weight $\lambda \geq 0$.
- The regularizer $R(\boldsymbol{x})$ can be defined such that it penalizes large variations in $\boldsymbol{x}$.

Image Prior and Regularization Term

Mathematical framework for image priors in super-resolution:

- We define the prior distribution according to the exponential form:

$$p(\boldsymbol{x}) = \frac{1}{Z} \exp\left(-\lambda R(\boldsymbol{x})\right),$$

  where Z is a normalization constant, and $R(\boldsymbol{x})$ denotes a regularization term with regularization weight $\lambda \geq 0$.
- The regularizer $R(\boldsymbol{x})$ can be defined such that it penalizes large variations in $\boldsymbol{x}$.
- The regularization weight λ measures the impact of the prior:
  - $\lambda = 0$: simple ML estimation,
  - $\lambda \to \infty$: estimation dominated by the prior distribution.

Selected Priors for Super-Resolution

Gaussian prior (Tikhonov regularization):

- $p(\mathbf{x})$ is a normal distribution, i. e., $\boldsymbol{x} \sim N(0, \boldsymbol{\Sigma})$ and accordingly:

$$p(\boldsymbol{x}) = \frac{1}{(2\pi)^{N/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}\mathbf{x}^T\boldsymbol{\Sigma}^{-1}\mathbf{x}\right).$$

Selected Priors for Super-Resolution

Gaussian prior (Tikhonov regularization):

- $p(\mathbf{x})$ is a normal distribution, i. e., $\boldsymbol{x} \sim N(0, \boldsymbol{\Sigma})$ and accordingly:

$$p(\boldsymbol{x}) = \frac{1}{(2\pi)^{N/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}\mathbf{x}^T\boldsymbol{\Sigma}^{-1}\mathbf{x}\right).$$

- Then, MAP estimation can be expressed as:

$$\hat{\mathbf{x}}_{\text{MAP}} = \arg\min_{\mathbf{x}} ||\mathbf{y} - \mathbf{W}\mathbf{x}||_2^2 + \lambda\mathbf{x}^\top\boldsymbol{\Sigma}^{-1}\mathbf{x} = \arg\min_{\mathbf{x}} ||\mathbf{y} - \mathbf{W}\mathbf{x}||_2^2 + \lambda\,||\boldsymbol{Q}\mathbf{x}||_2^2.$$

  $\boldsymbol{Q}\mathbf{x}$ computes variations in $\boldsymbol{x}$, e. g., gradient (first-order) or curvature (second-order).

Selected Priors for Super-Resolution

Gaussian prior (Tikhonov regularization):

- $p(\mathbf{x})$ is a normal distribution, i. e., $\boldsymbol{x} \sim N(0, \boldsymbol{\Sigma})$ and accordingly:

$$p(\boldsymbol{x}) = \frac{1}{(2\pi)^{N/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}\mathbf{x}^T\boldsymbol{\Sigma}^{-1}\mathbf{x}\right).$$

- Then, MAP estimation can be expressed as:

$$\hat{\mathbf{x}}_{\text{MAP}} = \arg\min_{\mathbf{x}} ||\mathbf{y} - \mathbf{W}\mathbf{x}||_2^2 + \lambda \mathbf{x}^\top \boldsymbol{\Sigma}^{-1}\mathbf{x} = \arg\min_{\mathbf{x}} ||\mathbf{y} - \mathbf{W}\mathbf{x}||_2^2 + \lambda\, ||\boldsymbol{Q}\mathbf{x}||_2^2.$$

  $\boldsymbol{Q}\mathbf{x}$ computes variations in $\boldsymbol{x}$, e. g., gradient (first-order) or curvature (second-order).

- This regularization favors smooth solutions for $\boldsymbol{x}$.

Selected Priors for Super-Resolution

Gaussian prior (Tikhonov regularization):

- $p(\mathbf{x})$ is a normal distribution, i. e., $\boldsymbol{x} \sim N(0, \boldsymbol{\Sigma})$ and accordingly:

$$p(\boldsymbol{x}) = \frac{1}{(2\pi)^{N/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}\mathbf{x}^T\boldsymbol{\Sigma}^{-1}\mathbf{x}\right).$$

- Then, MAP estimation can be expressed as:

$$\hat{\mathbf{x}}_{\mathrm{MAP}} = \arg\min_{\mathbf{x}} ||\mathbf{y} - \mathbf{W}\mathbf{x}||_2^2 + \lambda \mathbf{x}^\top \boldsymbol{\Sigma}^{-1}\mathbf{x} = \arg\min_{\mathbf{x}} ||\mathbf{y} - \mathbf{W}\mathbf{x}||_2^2 + \lambda\, ||\boldsymbol{Q}\mathbf{x}||_2^2.$$

  $\boldsymbol{Q}\mathbf{x}$ computes variations in $\boldsymbol{x}$, e. g., gradient (first-order) or curvature (second-order).
- This regularization favors smooth solutions for $\boldsymbol{x}$.
- It facilitates a closed-form solution, but does not preserve discontinuities.

Selected Priors for Super-Resolution

Total variation (TV):

- $p(\boldsymbol{x})$ is modeled by a Laplacian distribution:

$$p(\boldsymbol{x}) \propto \exp\left(-\lambda\,||\boldsymbol{Qx}||_1\right).$$

Selected Priors for Super-Resolution

Total variation (TV):

- $p(\boldsymbol{x})$ is modeled by a Laplacian distribution:

$$p(\boldsymbol{x}) \propto \exp\left(-\lambda \, ||\boldsymbol{Q}\mathbf{x}||_1\right).$$

- We define an isotropic version of the TV prior:

$$p(\boldsymbol{x}) \propto \exp\left(-\lambda \, R(\boldsymbol{x})\right), \quad \text{where } R(\boldsymbol{x}) = \sqrt{||\boldsymbol{Q}_u\boldsymbol{x}||_2^2 + ||\boldsymbol{Q}_v\boldsymbol{x}||_2^2}.$$

$\boldsymbol{Q}_u$ and $\boldsymbol{Q}_v$ are gradient operators for horizontal and vertical direction, respectively.

Selected Priors for Super-Resolution

Total variation (TV):

- $p(\boldsymbol{x})$ is modeled by a Laplacian distribution:

$$p(\boldsymbol{x}) \propto \exp\left(-\lambda \, ||\boldsymbol{Q}\mathbf{x}||_1\right).$$

- We define an isotropic version of the TV prior:

$$p(\boldsymbol{x}) \propto \exp\left(-\lambda R(\boldsymbol{x})\right), \quad \text{where } R(\boldsymbol{x}) = \sqrt{||\boldsymbol{Q}_u\boldsymbol{x}||_2^2 + ||\boldsymbol{Q}_v\boldsymbol{x}||_2^2}.$$

  $\boldsymbol{Q}_u$ and $\boldsymbol{Q}_v$ are gradient operators for horizontal and vertical direction, respectively.

- Then, MAP estimation can be expressed as:

$$\hat{\mathbf{x}}_{\text{MAP}} = \arg\min_{\mathbf{x}} ||\mathbf{y} - \mathbf{W}\mathbf{x}||_2^2 + \lambda \sqrt{||\boldsymbol{Q}_u\boldsymbol{x}||_2^2 + ||\boldsymbol{Q}_v\boldsymbol{x}||_2^2}.$$

Selected Priors for Super-Resolution

Total variation (TV):

- $p(\boldsymbol{x})$ is modeled by a Laplacian distribution:

$$p(\boldsymbol{x}) \propto \exp\left(-\lambda \, ||\boldsymbol{Q}\mathbf{x}||_1\right).$$

- We define an isotropic version of the TV prior:

$$p(\boldsymbol{x}) \propto \exp\left(-\lambda \, R(\boldsymbol{x})\right), \quad \text{where } R(\boldsymbol{x}) = \sqrt{||\boldsymbol{Q}_u\boldsymbol{x}||_2^2 + ||\boldsymbol{Q}_v\boldsymbol{x}||_2^2}.$$

  $\boldsymbol{Q}_u$ and $\boldsymbol{Q}_v$ are gradient operators for horizontal and vertical direction, respectively.
- Then, MAP estimation can be expressed as:

$$\hat{\mathbf{x}}_{\text{MAP}} = \arg\min_{\mathbf{x}} ||\mathbf{y} - \mathbf{W}\mathbf{x}||_2^2 + \lambda \sqrt{||\boldsymbol{Q}_u\boldsymbol{x}||_2^2 + ||\boldsymbol{Q}_v\boldsymbol{x}||_2^2}.$$

- It favors piecewise constant solutions for $\boldsymbol{x}$ (edge preserving regularization).

Selected Priors for Super-Resolution

Bilateral total variation (BTV):

- BTV is a generalization of the TV prior

Selected Priors for Super-Resolution

Bilateral total variation (BTV):

- BTV is a generalization of the TV prior
- Compare **x** to multiple shifted versions as generalization of the gradient in a $(2P+1) \times (2P+1)$ window:

$$p(\mathbf{x}) \propto \exp\left(-\lambda \sum_{u=-P}^{P} \sum_{v=-P}^{P} \alpha^{|u|+|v|} \left|\left|\mathbf{x} - \mathbf{S}_i^u \mathbf{S}_j^v \mathbf{x}\right|\right|_1\right).$$

$\mathbf{S}_i^u$ and $\mathbf{S}_j^v$ are shifts of u and v pixels in horizontal and vertical direction, respectively.

Selected Priors for Super-Resolution

Bilateral total variation (BTV):

- BTV is a generalization of the TV prior
- Compare **x** to multiple shifted versions as generalization of the gradient in a $(2P+1)\times(2P+1)$ window:

$$p(\mathbf{x}) \propto \exp\left(-\lambda \sum_{u=-P}^{P} \sum_{v=-P}^{P} \alpha^{|u|+|v|} \left|\left|\mathbf{x} - \mathbf{S}_i^u \mathbf{S}_j^v \mathbf{x}\right|\right|_1\right).$$

  $\mathbf{S}_i^u$ and $\mathbf{S}_j^v$ are shifts of u and v pixels in horizontal and vertical direction, respectively.
- Then, MAP estimation can be expressed as:

$$\hat{\mathbf{x}}_{\text{MAP}} = \arg\min_{\mathbf{x}} ||\mathbf{y} - \mathbf{W}\mathbf{x}||_2^2 + \lambda \left(\sum_{u,v=-n}^{n} \alpha^{|u|+|v|} \left|\left|\mathbf{x} - \mathbf{S}_i^u \mathbf{S}_j^v \mathbf{x}\right|\right|_1\right).$$

Selected Priors for Super-Resolution

Bilateral total variation (BTV):

- BTV is a generalization of the TV prior
- Compare **x** to multiple shifted versions as generalization of the gradient in a $(2P+1)\times(2P+1)$ window:

$$p(\mathbf{x}) \propto \exp\left(-\lambda \sum_{u=-P}^{P} \sum_{v=-P}^{P} \alpha^{|u|+|v|} \left|\left|\mathbf{x}-\mathbf{S}_i^u \mathbf{S}_j^v \mathbf{x}\right|\right|_1\right).$$

  $\mathbf{S}_i^u$ and $\mathbf{S}_j^v$ are shifts of u and v pixels in horizontal and vertical direction, respectively.
- Then, MAP estimation can be expressed as:

$$\hat{\mathbf{x}}_{\mathrm{MAP}} = \arg\min_{\mathbf{x}} ||\mathbf{y}-\mathbf{W}\mathbf{x}||_2^2 + \lambda\left(\sum_{u,v=-n}^{n} \alpha^{|u|+|v|} \left|\left|\mathbf{x}-\mathbf{S}_i^u \mathbf{S}_j^v \mathbf{x}\right|\right|_1\right).$$

- It reduces "‘staircasing"’ artifacts which occur using the common TV prior.

Selected Priors for Super-Resolution

Huber prior:

- $p(\boldsymbol{x})$ is modeled by the Huber loss function:

$$p(\boldsymbol{x}) \propto \exp\left(-\lambda \sum_{i=1}^{N} h_\tau([\boldsymbol{Qx}]_i)\right),$$

$$h_\tau(z) = \begin{cases} \frac{1}{2}z^2, & \text{if } |z| \leq \tau, \\ \tau\left(|z| - \frac{\tau}{2}\right), & \text{otherwise.} \end{cases}$$

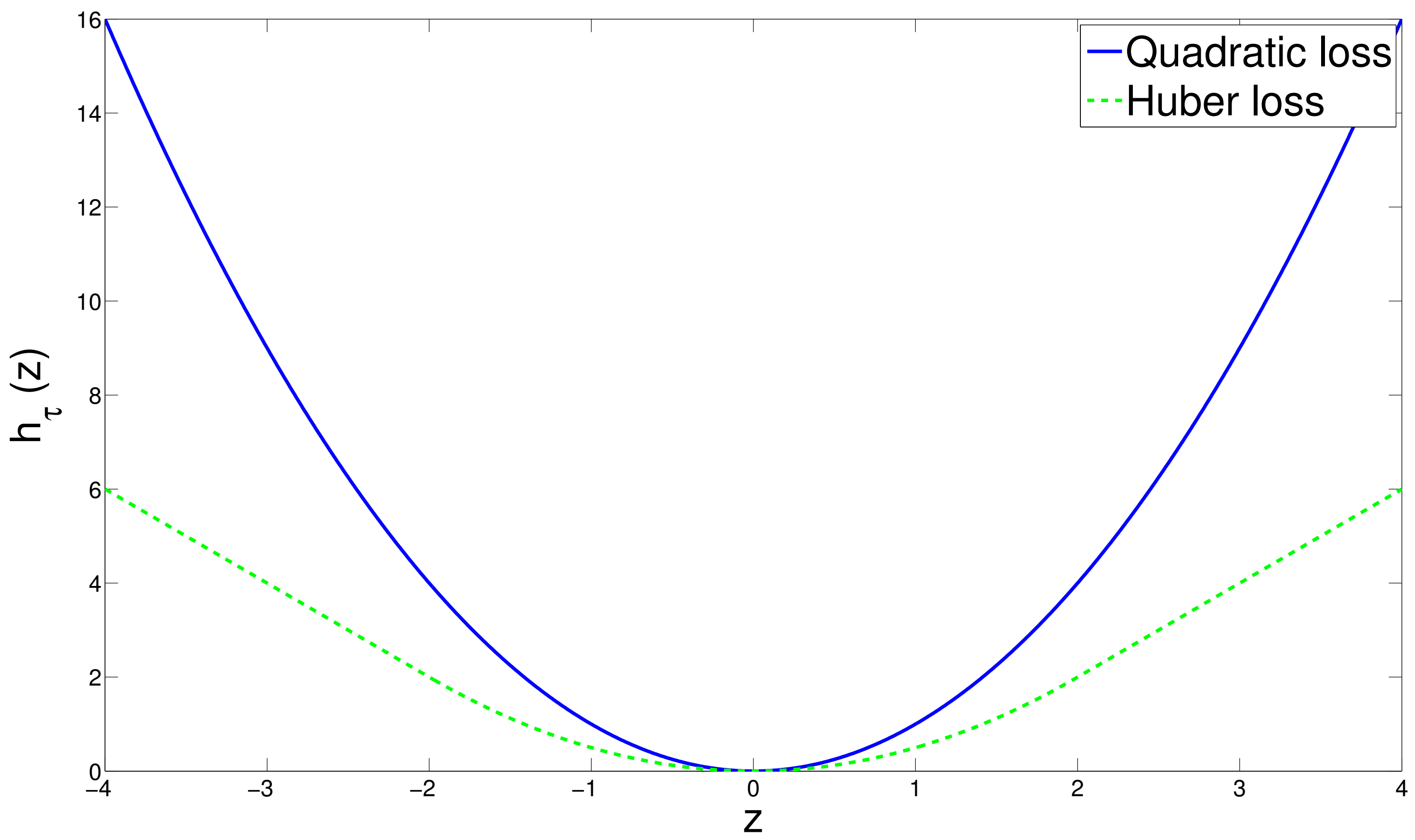

Figure 2: Comparison of quadratic loss and Huber loss

Selected Priors for Super-Resolution

Huber prior:

- $p(\boldsymbol{x})$ is modeled by the Huber loss function:

$$p(\boldsymbol{x}) \propto \exp\left(-\lambda \sum_{i=1}^{N} h_\tau([\mathbf{Qx}]_i)\right),$$

$$h_\tau(z) = \begin{cases} \frac{1}{2}z^2, & \text{if } |z| \leq \tau, \\ \tau\left(|z| - \frac{\tau}{2}\right), & \text{otherwise.} \end{cases}$$

- Then, MAP estimation can be expressed as:

$$\arg\min_{\mathbf{x}} ||\mathbf{y} - \mathbf{Wx}||_2^2 + \lambda \sum_{i=1}^{N} h_\tau([\mathbf{Qx}]_i).$$

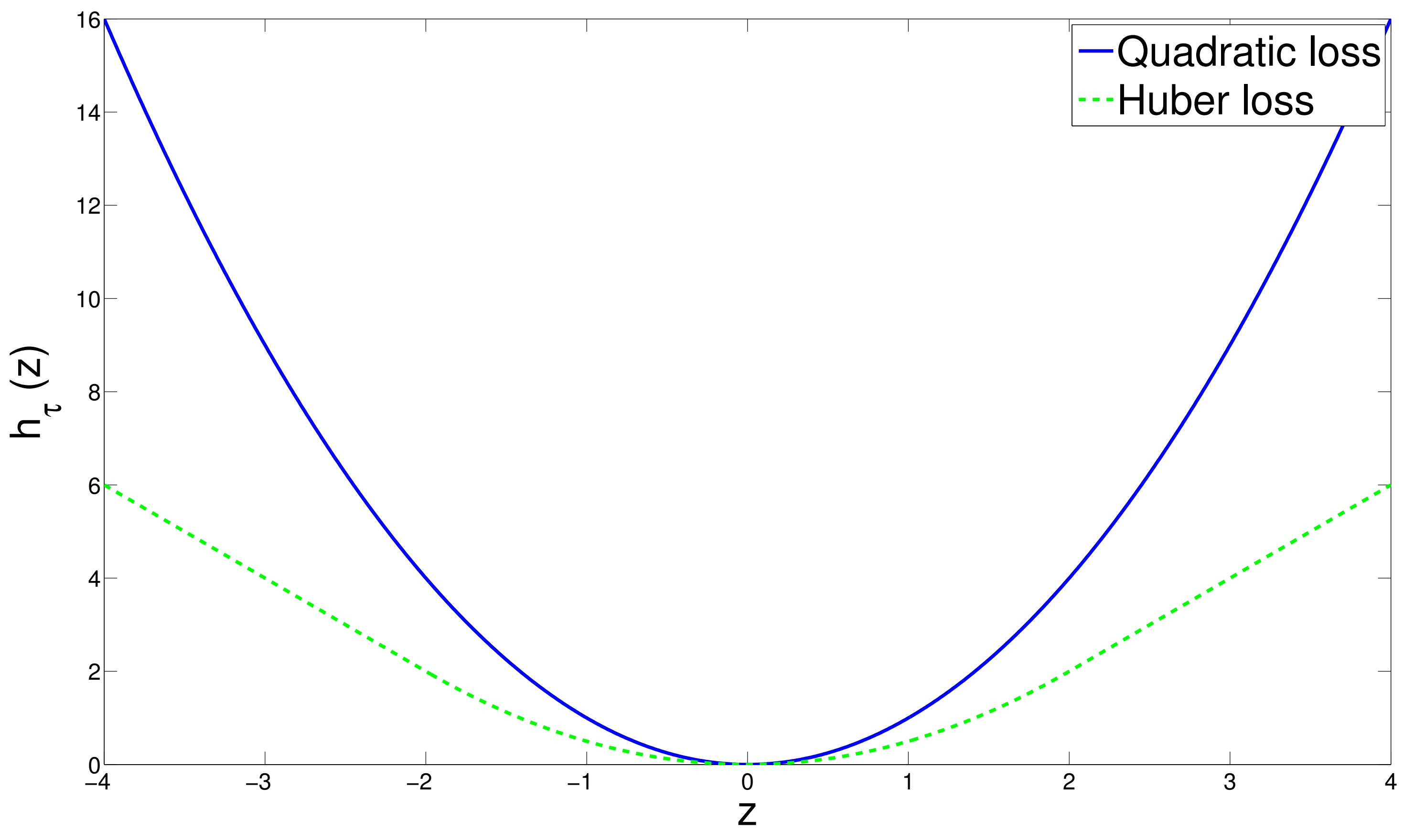

Figure 2: Comparison of quadratic loss and Huber loss

Selected Priors for Super-Resolution

Huber prior:

- $p(\boldsymbol{x})$ is modeled by the Huber loss function:

$$p(\boldsymbol{x}) \propto \exp\left(-\lambda \sum_{i=1}^{N} h_\tau([\boldsymbol{Q}\mathbf{x}]_i)\right),$$

$$h_\tau(z) = \begin{cases} \frac{1}{2}z^2, & \text{if } |z| \leq \tau, \\ \tau\left(|z| - \frac{\tau}{2}\right), & \text{otherwise.} \end{cases}$$

- Then, MAP estimation can be expressed as:

$$\arg\min_{\mathbf{x}} ||\mathbf{y} - \mathbf{W}\mathbf{x}||_2^2 + \lambda \sum_{i=1}^{N} h_\tau([\mathbf{Q}\mathbf{x}]_i).$$

- It yields piecewise smooth solutions (edge preserving regularization).

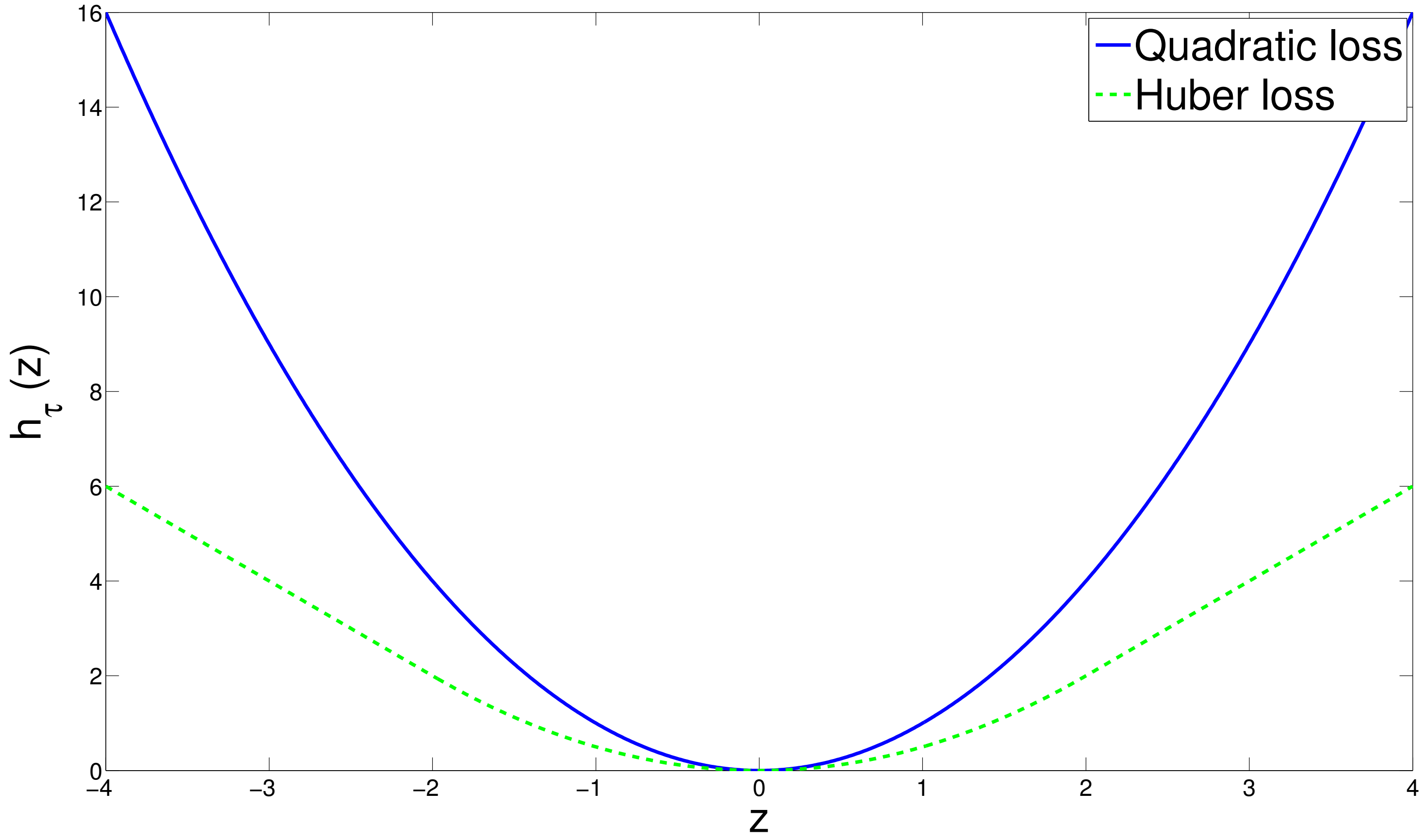

Figure 2: Comparison of quadratic loss and Huber loss

ML and MAP Estimation: Discussion

Figure 3: Low-resolution frame (left), ML estimation (middle), MAP estimation with TV prior (right)

Topics

Maximum A Posteriori Estimation

Image Priors

Summary

Take Home Messages

Further Readings

Take Home Messages

- Super-resolution is an ill-posed problem, so using regularization by utilizing prior knowledge is a good idea.
- We introduced several regularizers: Tikhonov, TV, BTV, and the edge-preserving approach using the Huber loss function.
- Regularization needs extra thought when implementing, but it is an important method to achieve adequate image quality.

Further Readings

Theory of image super-resolution (books and review articles):

- Hayit Greenspan. "Super-Resolution in Medical Imaging". In: *The Computer Journal* 52.1 (Feb. 2008), pp. 43–63. DOI: 10.1093/comjnl/bxm075
- Peyman Milanfar, ed. *Super-Resolution Imaging*. Digital Imaging and Computer Vision. CRC Press, 2011
- Sina Farsiu et al. "Advances and Challenges in Super-Resolution". In: *International Journal of Imaging Systems and Technology* 14.2 (Aug. 2004), pp. 47–57. DOI: 10.1002/ima.20007
- Sung Cheol Park, Min Kyu Park, and Moon Gi Kang. "Super-Resolution Image Reconstruction: A Technical Overview". In: *IEEE Signal Processing Magazine* 20.3 (May 2003), pp. 21–36. DOI: 10.1109/MSP.2003.1203207

ML/MAP super-resolution:

- Lyndsey C. Pickup. "Machine Learning in Multi-frame Image Super-resolution". PhD Thesis. Robotics Research Group, University of Oxford, 2007
- Michael Elad and Arie Feuer. "Restoration of a Single Superresolution Image from Several Blurred, Noisy, and Undersampled Measured Images". In: *IEEE Transactions on Image Processing* 6.12 (Dec. 1997), pp. 1646–1658. DOI: 10.1109/83.650118