# 1 Bayesian Image Denoising



$$P(\text{NoiselessImage}|\text{NoisyImage}) = \frac{P(\text{NoisyImage}|\text{NoiselessImage})P(\text{NoiselessImage})}{P(\text{NoisyImage})}$$

– **Likelihood** PDF = **noise** model = probability of generating the data given the noiseless image

– **Prior** PDF = our prior beliefs about the noiseless image **before** observing the data

– **Posterior** PDF: product of likelihood and prior (with a normalization factor in the denominator to make it a PDF)
Called "posterior" because it is what we get **"post"** / **after** observing the data

– Goal: We want to find the noiseless image that maximizes the posterior !

# 2 Denoising Optimization via Iterated Conditional Mode (ICM)

– Consider noiseless image modeled as a MRF $X = x$, with, say, real-valued intensities
– Consider observed image data $Y = y$
– Noise Model for Intensities Given Noiseless Intensities : $P(Y|X) := \Pi_i P(Y_i|X_i)$
e.g., assuming that the noise is additive i.i.d. zero-mean Gaussian, $P(Y_i|X_i) = G(y_i|x_i, \sigma^2)$
– Let $\theta =$ parameters underlying the noise model and the MRF model (in general)

– **Optimization Problem and Strategy :**
(1) Assume that the MRF parameters are user controlled
(2) Assume that the noise level is estimated using the ML estimate in the background region, where signal is known to be zero
(3) Obtain the noiseless image $x$ by maximizing $\max_{x,\theta} P(x|y,\theta)$

– **Obtain MAP estimate for Noiseless Image:** $\max_x P(x|y,\theta)$
– Rewrite the objective function as

$$\begin{aligned} P(X|y,\theta) &= P(X_i, X_{\sim i}|y,\theta) \\ &= P(X_i|X_{\sim i}, y, \theta)P(X_{\sim i}|y,\theta) \text{ Conditional Probability} \\ &= P(X_i|X_{N_i}, y, \theta)P(X_{\sim i}|y,\theta) \text{ Markov assumption on } X \\ &= P(X_i|X_{N_i}, y_i, \theta)P(X_{\sim i}|y,\theta) \text{ Conditional independence assumption in noise model } P(Y|X) \end{aligned}$$

– For ICM optimization, at a chosen voxel $i$, perform

$$\begin{aligned}\max_{x_i} P(X|y,\theta) &= \max_{x_i} P(X_i|X_{N_i}, y_i, \theta) P(X_{\sim i}|y,\theta) \\ &= \max_{x_i} P(X_i|X_{N_i}, y_i, \theta) \text{ Second term doesn't depend on } x_i \\ &= \max_{x_i} \frac{P(y_i|X_i, X_{N_i}, \theta) P(X_i|X_{N_i}, \theta)}{P(y_i|X_{N_i}, \theta)} \text{ Bayes Rule} \\ &= \max_{x_i} P(y_i|X_i, X_{N_i}, \theta) P(X_i|X_{N_i}, \theta) \text{ Denominator doesn't depend on } x_i \\ &= \max_{x_i} P(y_i|X_i, \theta) P(X_i|X_{N_i}, \theta) \text{ Conditional independence assumption in noise model } P(Y|X)\end{aligned}$$

where
– the first term $P(y_i|X_i, \theta)$ is the likelihood function = noise model
– the second term $P(X_i|X_{N_i}, \theta)$ is the local/conditional prior on the noiseless image
– This optimization seeks the **mode** of the **local/conditional posterior** !! Hence, the name ICM.

– **Order of Intensity Updates:**
If we want every update to necessarily increase the posterior $P(x|y,\theta)$, then
(1) Sequentially: Column by column, and then row by row (but, may lead to artifacts)
(2) Sequentially: Randomized order each iteration (need to generate random sequence each iteration. Are artifacts eliminated ?)
(3) In Parallel: If **maximizing**, doesn't guarantee increase in posterior probability (**note: NOT applicable to gradient descent**). No need to generate random sequence. Are artifacts eliminated ?
(4) In Parallel: **No need to seek the mode; just go towards the mode**. Gradient ascent/descent with dynamic step size and objective-function monitoring. Guarantees increase in posterior probability. No need to generate random sequence. Are artifacts eliminated ? Is the solution reached faster / slower ?

– The local conditional prior on the noiseless image is

$$P(X_i|X_{N_i}, \theta) = \frac{1}{Z_i} \exp\left(-\sum_{a \in A} V_a(x_a)\right) \text{ where } A \text{ is the set of cliques that contain site } i$$

# 3 Fully-Sampled MRI (Complex) (Noise: Gaussian)

The **circularly-symmetric uni-variate** Gaussian noise model on **complex numbers** is :

$$P(y|x) = \Pi_i P(y_i|x_i) = \Pi_i G_{\mathbb{C}}(y_i|x_i, \sigma^2) = \Pi_i \frac{1}{\sigma^2 \pi} \exp\left(-\frac{|y_i - x_i|^2}{\sigma^2}\right)$$

– Note: the general (asymmetric) multivariate complex Gaussian PDF is more complex with 3 parameters: mean vector, (Hermitian + non-negative definite) covariance matrix, and **(symmetric) relation matrix** ! See Wikipedia.

For **ICM optimization**, at a chosen voxel $i$, perform

$$\begin{aligned}
\max_{x_i} P(x|y,\theta) &= \max_{x_i} P(y_i|x_i,\theta) P(x_i|x_{N_i},\theta) \\
&= \max_{x_i} \Big( \log P(y_i|x_i,\theta) + \log P(x_i|x_{N_i},\theta) \Big) \\
&= \max_{x_i} \left( \frac{-|y_i - x_i|^2}{\sigma^2} + \sum_{a \in A_i} (-V_a(x_a)) \right) \\
&= \min_{x_i} \left( \frac{|y_i - x_i|^2}{\sigma^2} + \sum_{a \in A_i} V_a(x_a) \right)
\end{aligned}$$

**Fidelity term:** $|y_i - x_i|^2/\sigma^2$ penalizes the deviation (infidelity) of the estimate $x$ from the data $y$

**Regularity term:** $\sum_{a \in A_i} V_a(x_a)$ penalizes the irregularity / roughness of the estimate $x$

For **gradient-descent optimization:** , at a chosen voxel $i$, the derivative is

$$\frac{\partial P(x|y,\theta)}{\partial x_i} = \frac{1}{\sigma^2} 2(x_i - y_i) + \frac{\partial}{\partial x_i} \sum_{a \in A_i} V_a(x_a)$$

Note: Fidelity term in the objective function is a real-valued function of complex variables. So, use the real derivative !!
For the entire image $x$, the gradient (column vector) is

$$g_2(x) := \left( \cdots, \frac{\partial P(x|y,\theta)}{\partial x_i}, \cdots \right)$$

Current solution $x^n$ at iteration $n$. Stepsize $\tau$. Updated solution is :

$$x^{n+1} = x^n - \tau g(x)$$

# 4 MRI Magnitude Images (Noise: Rician)

– The observed noisy (magnitude-MR) image data $y$ is real
– The noiseless (magnitude-MR) image $x$ is real

– The **prior PDF** remains the same

The **likelihood PDF** is

$$P(y|x) := \Pi_i P(y_i|x_i) := \Pi_i \frac{y_i}{\sigma^2} \exp\left(-\frac{y_i^2 + x_i^2}{2\sigma^2}\right) I_0\left(\frac{y_i x_i}{\sigma^2}\right)$$

where $I_0(z)$ is the modified Bessel function of the first kind with order zero

For **ICM optimization**, at a chosen voxel $i$, perform

$$\max_{x_i} P(x|y,\theta) = \min_{x_i} \left(\frac{y_i^2 + x_i^2}{2\sigma^2} - \log I_0\left(\frac{y_i x_i}{\sigma^2}\right) + \sum_{a \in A_i} V_a(x_a)\right)$$

For **gradient-descent optimization:** , at a chosen voxel $i$, the derivative is

$$\frac{\partial P(x|y,\theta)}{\partial x_i} = \frac{x_i}{\sigma^2} - \frac{I_1\left(\frac{y_i x_i}{\sigma^2}\right)}{I_0\left(\frac{y_i x_i}{\sigma^2}\right)} \frac{y_i}{\sigma^2} + \frac{\partial}{\partial x_i} \sum_{a \in A_i} V_a(x_a)$$

where $I_1(z)$ is the modified Bessel function of the first kind with order **one** !! $\partial I_0(z)/\partial z = I_1(z)$

# 5 Ultrasound Magnitude Images (Noise: Speckle)

– The observed noisy image data $y$ is real
– The noiseless image $x$ is real

– The **prior PDF** remains the same

– The speckle-noise model is $Y = X + \sqrt{X}Z$ where $P(Z) := G(0, \sigma^2)$. This implies

For the RV $Z, P(z) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{z^2}{2\sigma^2}\right)$

For the transformed RV $Y_1 := \sqrt{x}Z, P(y_1) = \frac{1}{\sqrt{x}\sigma\sqrt{2\pi}} \exp\left(-\frac{y_1^2}{2x\sigma^2}\right)$

Transformation maintains PDF as Gaussian and scales std. dev. by $\sqrt{x}$

For the transformed RV $Y_2 := Y_1 + x, P(y_2) = \frac{1}{\sqrt{x}\sigma\sqrt{2\pi}} \exp\left(-\frac{(y_2 - x)^2}{2x\sigma^2}\right)$

Transformation maintains PDF as Gaussian and translates mean by $x$

The **likelihood PDF** is

$$P(y|x) := \Pi_i P(y_i|x_i) := \Pi_i G(y_i|x_i, x_i\sigma^2) = \frac{1}{\sqrt{x}\sigma\sqrt{2\pi}} \exp\left(-\frac{(y_i - x_i)^2}{2x_i\sigma^2}\right)$$

For **ICM optimization**, at a chosen voxel $i$, perform

$$\max_{x_i} P(x|y,\theta) = \min_{x_i} \left(\frac{\log(x_i)}{2} + \frac{(y_i - x_i)^2}{2x_i\sigma^2} + \sum_{a \in A_i} V_a(x_a)\right)$$

For **gradient-descent optimization**, at a chosen voxel $i$, the derivative is

$$\frac{\partial P(x|y,\theta)}{\partial x_i} = \frac{1}{2x_i} + \frac{1}{2\sigma^2}\left(\frac{x_i^2 - y_i^2}{x_i^2}\right) + \frac{\partial}{\partial x_i} \sum_{a \in A_i} V_a(x_a)$$

# 6 Weighting Likelihood and MRF Prior

– **Motivation :**
(1) We want to **control** the strength of the **prior** model based on certain criteria (e.g., noise level)
(2) We want to **balance** the enforcement of **fidelity and regularity** based on certain criteria (e.g., noise level)

– **Thumb Rule :**
(1) Very high noise levels $\implies$ data is corrupted $\implies$ use a strong prior
(2) Very low (or zero) noise levels $\implies$ data is high quality $\implies$ use a weak prior

– **How to do it ?**
– Introduce a user-controlled parameter $\beta \in [0, 1]$ that defines the balance between (1) enforcing the prior model and (2) enforcing the likelihood model

## 6.1 Weighting MRF Prior

In the **prior** PDF, introduce a parameter $\beta \in [0, 1]$ s.t.

$P(x) := \frac{1}{Z(\beta)} \exp\left(-\beta \frac{1}{T} U(x)\right)$ where

$U(x) := \sum_{c \in C} V_c(x_c)$ where

$Z(\beta) := \sum_{x} \exp\left(-\beta \frac{1}{T} U(x)\right)$

This changes the **local conditional prior** to

$P(x_i | x_{N_i}, \theta) = \frac{1}{Z_i(\beta)} \exp\left(-\beta \frac{1}{T} \sum_{a \in A} V_a(x_a)\right)$ where $A$ is the set of cliques that contain site $i$

**Interpretation:** Introducing $\beta$ is similar to changing the temperature $T$

## 6.2 Weighting Likelihood (Complex-Gaussian Noise)

In the **likelihood** PDF (Complex Gaussian), introduce a parameter $\alpha \in [0, 1]$, where $\alpha := 1 - \beta$ s.t.

$$P(y|x) := \Pi_i P(y_i|x_i) := \Pi_i G_\alpha(y_i|x_i, \sigma^2) \text{ where}$$

$$G_\alpha(y_i|x_i, \sigma^2) := \frac{1}{Z(\sigma, \alpha)} \exp\left(-\alpha \frac{|y_i - x_i|^2}{\sigma^2}\right) \text{ where}$$

$$Z(\sigma, \alpha) := \int_{y=-\infty}^{\infty} \exp\left(-\alpha \frac{|y - x_i|^2}{\sigma^2}\right) dx = \int_y \exp\left(-\frac{|y - x_i|^2}{(\sigma/\sqrt{\alpha})^2}\right) dx = \frac{1}{(\sigma/\sqrt{\alpha})^2 \pi}$$

**Interpretation:** Introducing $\alpha$ is similar to changing the noise level / standard deviation $\sigma$
This is the Complex-Gaussian PDF with parameters $(x_i, \sigma^2/\alpha)$

## 6.3 Modified Optimization Problem (Complex-Gaussian Noise)

For **ICM optimization**, at a chosen voxel $i$, perform

$$\max_{x_i} P(x|y, \theta) = \min_{x_i} \left( (1 - \beta) \frac{|y_i - x_i|^2}{\sigma^2} + \beta \sum_{a \in A_i} V_a(x_a) \right)$$

For **gradient-descent optimization**, at a chosen voxel $i$, the derivative is

$$\frac{\partial P(x|y, \theta)}{\partial x_i} = (1 - \beta) \frac{1}{\sigma^2} 2(x_i - y_i) + \beta \frac{\partial}{\partial x_i} \sum_{a \in A_i} V_a(x_a)$$

**Interpretation:**

$\beta = 0$ ignores the prior and we get the ML estimate.
Note: $\beta = 0$ is like temperature $T \to \infty$, making all images equally likely !!

$\beta = 1$ ignores the data and we may get an oversmooth solution.
Note: $\beta = 1$ implies $\alpha = 0$, which is like $\sigma \to \infty$, making the likelihood a uniform PDF !!

### 6.4 Weighting Likelihood (Rician Noise)

In the **likelihood** PDF (Complex Gaussian), introduce a parameter $\alpha \in [0, 1]$, where $\alpha := 1 - \beta$ s.t.

$$P(y|x) := \Pi_i P(y_i|x_i) := \Pi_i \frac{\alpha y_i}{\sigma^2} \exp\left(-\alpha \frac{y_i^2 + x_i^2}{2\sigma^2}\right) I_0\left(\alpha \frac{y_i x_i}{\sigma^2}\right)$$

**Interpretation:** Introducing $\alpha$ is similar to changing the noise level $\sigma$
This is the Rician PDF with parameters $(x_i, \sigma/\sqrt{\alpha})$

### 6.5 Modified Optimization Problem (Rician Noise)

For **ICM optimization**, at a chosen voxel $i$, perform

$$\max_{x_i} P(x|y,\theta) = \min_{x_i} \left( (1-\beta)\frac{y_i^2 + x_i^2}{2\sigma^2} - \log I_0\left((1-\beta)\frac{y_i x_i}{\sigma^2}\right) + \beta \sum_{a \in A_i} V_a(x_a) \right)$$

For **gradient-descent optimization**, at a chosen voxel $i$, the derivative is

$$\frac{\partial P(x|y,\theta)}{\partial x_i} = (1-\beta)\frac{x_i}{\sigma^2} - \frac{I_1\left((1-\beta)\frac{y_i x_i}{\sigma^2}\right)}{I_0\left((1-\beta)\frac{y_i x_i}{\sigma^2}\right)}(1-\beta)\frac{y_i}{\sigma^2} + \beta \frac{\partial}{\partial x_i} \sum_{a \in A_i} V_a(x_a)$$

where $I_1(z)$ is the modified Bessel function of the first kind with order **one** !! $\partial I_0(z)/\partial z = I_1(z)$

**Interpretation:**

$\beta = 0$ ignores the prior and we get the ML estimate.
Note: $\beta = 0$ is like temperature $T \to \infty$, making all images equally likely !!

$\beta = 1$ ignores the data and we may get an oversmooth solution.
Note: $\beta = 1$ implies $\alpha = 0$, which is like $\sigma \to \infty$ (i.e., variance of noise in complex domain $\to \infty$), making the likelihood a uniform PDF !!

### 6.6 Weighting Likelihood (Speckle Noise)

In the **likelihood** PDF (Gaussian with signal-dependent variance), introduce a parameter $\alpha \in [0,1]$, where $\alpha := 1-\beta$ s.t.

$$P(y|x) := \Pi_i P(y_i|x_i) := \Pi_i G(y_i|x_i, x_i\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}\sqrt{x_i^\alpha/\alpha}} \exp\left(-\frac{(y_i-x_i)^2}{2(x_i^\alpha/\alpha)\sigma^2}\right)$$

**Interpretation:** Introducing $\alpha$ is similar to changing the noise level / standard deviation $\sigma$
This is the Gaussian PDF with parameters $(x_i, \sigma^2(x_i^\alpha/\alpha))$

### 6.7 Modified Optimization Problem (Speckle Noise)

For **ICM optimization**, at a chosen voxel $i$, perform

$$\max_{x_i} P(x|y,\theta) = \min_{x_i} \left(\alpha\frac{1}{2}\log(x_i) + \alpha\frac{1}{2\sigma^2}\frac{(y_i-x_i)^2}{x_i^\alpha} + \beta\sum_{a\in A_i} V_a(x_a)\right)$$

For **gradient-descent optimization**, at a chosen voxel $i$, the derivative is

$$\frac{\partial P(x|y,\theta)}{\partial x_i} = \alpha\frac{1}{2x_i} + \alpha\frac{1}{2\sigma^2}\frac{\partial}{\partial x_i}\frac{(y_i-x_i)^2}{x_i^\alpha} + \beta\frac{\partial}{\partial x_i}\sum_{a\in A_i} V_a(x_a)$$

**Interpretation:**

$\beta = 0$ ignores the prior and we get the ML estimate.
Note: $\beta = 0$ is like temperature $T \to \infty$, making all images equally likely !!

$\beta = 1 \implies \alpha = 0$ ignores the data and we may get an oversmooth solution
Note: $\beta = 1 \implies \alpha = 0$, which is like $\sigma \to \infty$ (i.e., variance of noise $\to \infty$), making the likelihood a uniform PDF !!