Lecture 16: Convex Potentials and the Scale Parameter

Chapter 6.3–6.4

Charles A. Bouman

Purdue University

Wednesday, September 30, 2026

Outline

1. **Convex potentials**: quadratic, L_1, GGMRF, Huber, QGGMRF
2. **The QGGMRF**: preserves edges and keeps the estimate stable
3. **The scale parameter** σ_x

Reading: Chapter 6.3–6.4, *Foundations of Computational Imaging*.

Today's Big Ideas

- A convex potential: stable MAP estimate and a global minimum
- The GMRF blurs edges; convex alternatives preserve them better
- The QGGMRF is quadratic at low contrast and grows like $|\Delta|^p$ at high contrast, with continuous first and second derivatives.
- A single scale parameter σ_x matches the model to the physical range of the data. For the GGMRF, its ML estimate has a closed form.

Convex potential functions

A potential function is **convex** if $\forall a, b \in \mathbb{R}$ and $\forall \lambda \in [0, 1]$,

$$\rho\left(\lambda a + (1-\lambda)b\right) \leq \lambda\rho(a) + (1-\lambda)\rho(b) \ .$$

When the data term is also convex, a convex potential has two major advantages.

- **Stable MAP estimate**: fewer undesirable artifacts
- The MAP problem is convex, so its **global maximum** is tractable

The simplest convex potential is the quadratic $\rho(\Delta) = \Delta^2$, which is the GMRF of Chapter 4. It is the most tractable, but its influence is unbounded, so it blurs sharp edges. The potentials below stay convex while preserving edges better.

The absolute value potential, or L_1 norm

One of the earliest alternatives is the absolute value potential of Besag, sometimes called an L_1 **norm** regularizer:

$$\rho(\Delta) = |\Delta| \ , \qquad \rho'(\Delta) = \text{sign}(\Delta) \ .$$

Here $\Delta = x_s - x_r$ is the difference between two neighboring pixels, and ρ' is the influence function of Lecture 15.

- It is **convex**.
- It grows more slowly than the quadratic, so it is **edge preserving**.
- It needs no threshold parameter T.

Notice that the influence is *constant* in magnitude. Every pixel difference, large or small, exerts the same force.

The L_1 estimate: flat regions and total variation

The L_1 cost is convex but not differentiable where $x_s = x_r$, and its minimum typically falls on those points. The estimate therefore contains groups of neighboring pixels of **equal value**: large flat regions with sharp transitions, sometimes described as **"waxy."**

- Because the cost is not continuously differentiable, gradient-based algorithms (gradient descent, conjugate gradient, ICD) are generally **not convergent**.

Relation to total variation

The L_1 potential is closely related to **total variation** regularization, $TV(x) = \int_{\mathbb{R}^2} |\nabla x(r)|\, dr$. Its discrete anisotropic form, $\sum_{i,j} |x_{i+1,j} - x_{i,j}| + |x_{i,j+1} - x_{i,j}|$, is exactly the pairwise Gibbs distribution with the $|\Delta|$ potential.

Generalizing L_1: the generalized Gaussian MRF (GGMRF)

The **generalized Gaussian MRF (GGMRF)** generalizes the L_1 norm to an L_p norm:

$$\rho(\Delta) = \frac{|\Delta|^p}{p} \ , \qquad \rho'(\Delta) = |\Delta|^{p-1}\text{sign}(\Delta) \ .$$

- For $p > 1$ it is **strictly convex and continuously differentiable**, so gradient-based algorithms reach a global minimum.
- No threshold T; only a **shape parameter** p, often 1.1–1.2

Setting $p = 1$ recovers the absolute value potential; $p = 2$ recovers the quadratic, up to a factor of $1/2$.

Rounding out L_1: the Huber function

The **Huber function** rounds out the point of the absolute value potential with a local quadratic near $\Delta = 0$:

$$\rho(\Delta) = \begin{cases} \frac{\Delta^2}{2} & \text{for } |\Delta| < T \ , \\ T|\Delta| - \frac{T^2}{2} & \text{for } |\Delta| \geq T \ , \end{cases} \qquad \rho'(\Delta) = \text{clip}\{\Delta, [-T, T]\} \ .$$

Advantage: near $\Delta = 0$ it keeps the properties of the GMRF, so it better maintains **low-contrast detail**.

Disadvantages: it needs a threshold T; for $|\Delta| > T$ it is not strictly convex; and its second derivative is discontinuous at $|\Delta| = T$. The $\log\cosh$ potential $\rho(\Delta) = T^2 \log\cosh(|\Delta|/T)$ fixes the last point with a continuous second derivative.

The Q-generalized GMRF (QGGMRF)

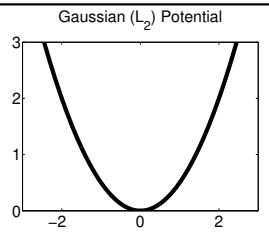
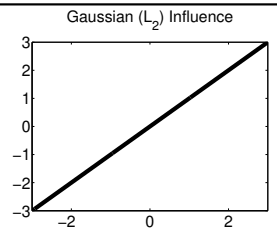
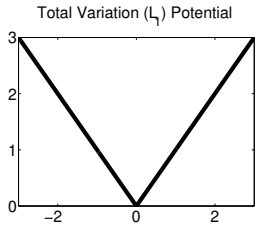
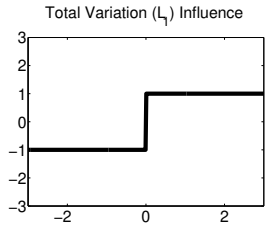
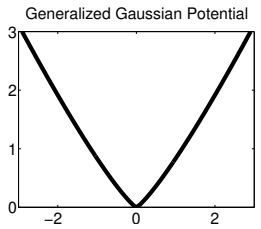
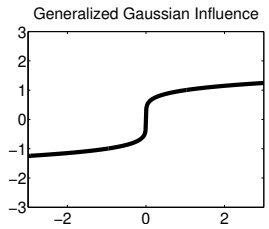
The **Q-generalized GMRF (QGGMRF)** controls the low-contrast behavior near $\Delta \approx 0$ and the high-contrast behavior as $|\Delta| \to \infty$ separately:

$$\rho(\Delta) = \frac{|\Delta|^p}{p} \left(\frac{\left|\frac{\Delta}{T}\right|^{q-p}}{1 + \left|\frac{\Delta}{T}\right|^{q-p}} \right) , \qquad 1 \leq p < q \ \text{ (convex).}$$

- High contrast, $|\Delta| > T$: $\rho(\Delta) \approx \frac{|\Delta|^p}{p}$, the GGMRF shape.
- Low contrast, $|\Delta| < T$: $\rho(\Delta) \approx \frac{|\Delta|^q}{p\, T^{q-p}}$.

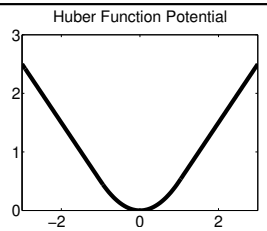
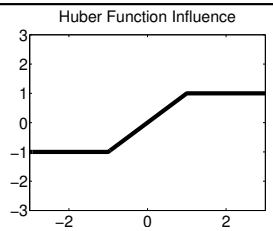
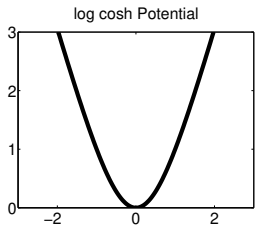
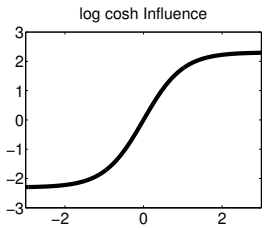
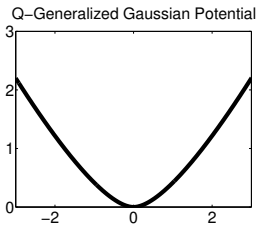
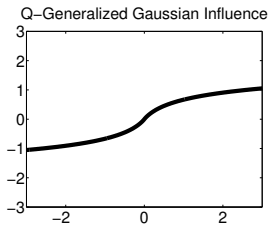
With $q = 2$ and $1 \leq p < q$ the potential is quadratic near $\Delta = 0$, with continuous first and second derivatives. This is not true of $|\Delta|^p$, whose second derivative is unbounded near $\Delta = 0$ for $1 \leq p < 2$; the continuous second derivative helps the majorization methods of Chapter 8. Lab 4 uses $p = 1.2$, $q = 2$.

Convex potentials: potential and influence (1 of 2)

	Potential $\rho(\Delta)$	Influence $\rho'(\Delta)$
Quadratic $\frac{\Delta^2}{2}$		
Total variation $\lvert\Delta\rvert$		
GGMRF $\frac{\lvert\Delta\rvert^p}{p}$		

All plots use $T = 1$, $p = 1.2$, $q = 2$.

Convex potentials: potential and influence (2 of 2)

	Potential $\rho(\Delta)$	Influence $\rho'(\Delta)$
Huber		
log cosh		
QGGMRF		

Each influence levels off, so a large difference across an edge exerts a bounded force.

The scale parameter σ_x

We almost always need to **scale the range of values** of an MRF model to the physical quantity it represents (image gray levels 0–255, ocean depth 0–11,000 m, and so on) without changing the basic character of the model.

We introduce a scaling parameter σ_x inside the potential functions of the Gibbs distribution:

$$p(x|\sigma_x) = \frac{1}{z} \exp\left\{ - \sum_{\{s,r\} \in \mathcal{P}} b_{s,r}\, \rho\left(\frac{x_s - x_r}{\sigma_x} \right) \right\} \ .$$

With this convention, if X has distribution $p(x|\sigma_x)$, then $\tilde{X} = X/\sigma_x$ has distribution $p(x|\sigma_x = 1)$.

ML estimation of σ_x

Estimating σ_x is more tractable than general MRF parameter estimation, because the partition function is a simple function of σ_x:

$$z(\sigma_x) = z_0 \, \sigma_x^N \ ,$$

with z_0 a constant. The ML estimate is then

$$\hat{\sigma}_x = \underset{\sigma_x > 0}{\arg\min} \left\{ u(x/\sigma_x) + N \log(\sigma_x) + \log(z_0) \right\} \ ,$$

where $u(x)$ is the energy function $u(x) = \sum_{\{s,r\} \in \mathcal{P}} b_{s,r} \, \rho(x_s - x_r)$. Setting the derivative to zero gives a single-variable rooting equation, solvable numerically for any image x:

$$\frac{\sigma_x}{N} \frac{d}{d\sigma_x} u(x/\sigma_x) \bigg|_{\sigma_x = \hat{\sigma}_x} = -1 \ .$$

No value of z_0 is needed, since z_0 does not depend on σ_x.

A closed form: the GGMRF case

When $u(\alpha x) = \alpha^p u(x)$, the ML estimate has the closed form $\hat{\sigma}_x = \left(\frac{p}{N} u(x)\right)^{1/p}$.

For the GGMRF, $\rho(\Delta) = |\Delta|^p/p$, the distribution is

$$p(x|\sigma_x) = \frac{1}{z} \exp\left\{ -\frac{1}{p\sigma_x^p} \sum_{\{s,r\}\in\mathcal{P}} b_{s,r}|x_s - x_r|^p \right\} ,$$

so the ML estimate of σ_x is

$$\hat{\sigma}_x^p = \frac{1}{N} \sum_{\{s,r\}\in\mathcal{P}} b_{s,r}|x_s - x_r|^p .$$

This is reminiscent of the variance of a Gaussian random variable, except that the sum is over all cliques and the differences are raised to the power p rather than squared.

Under- and over-regularization

Setting the single parameter σ_x is a major practical concern when $p(x)$ is used as a prior in MAP estimation.

σ_x too large: underregularized

The image overfits the measured data and has too much noisy variation.

σ_x too small: overregularized

Important detail is removed by oversmoothing.

For notational simplicity we fold σ_x into the potential. In the quadratic case,

$$\rho(\Delta) = \frac{\Delta^2}{2\sigma_x^2} \ , \qquad \rho'(\Delta) = \frac{\Delta}{\sigma_x^2} \ .$$

The transition point now sits at $\sigma_x T$: so T is a unitless shape parameter and σ_x carries the physical units.

The scaled convex potential functions

Name	$\rho(\Delta)$	$\rho'(\Delta)$
Quadratic	$\frac{\Delta^2}{2\sigma_x^2}$	$\frac{\Delta}{\sigma_x^2}$
Total Variation	$\frac{\lvert\Delta\rvert}{\sigma_x}$	$\frac{\text{sign}(\Delta)}{\sigma_x}$
Huber	$\begin{cases} \frac{\Delta^2}{2\sigma_x^2} & \text{for } \lvert\Delta\rvert < T\sigma_x \ , \\ T\left\lvert\frac{\Delta}{\sigma_x}\right\rvert - \frac{T^2}{2} & \text{for } \lvert\Delta\rvert \geq T\sigma_x \end{cases}$	$\frac{1}{\sigma_x^2}\text{clip}\{\Delta, [-T\sigma_x, T\sigma_x]\}$
log cosh	$T^2 \log\cosh\left(\frac{\Delta}{T\sigma_x}\right)$	$\frac{T}{\sigma_x}\tanh\left(\frac{\Delta}{T\sigma_x}\right)$
GGMRF	$\frac{\lvert\Delta\rvert^p}{p\sigma_x^p}$	$\frac{\lvert\Delta\rvert^{p-1}}{\sigma_x^p}\text{sign}(\Delta)$
QGGMRF	$\frac{\lvert\Delta\rvert^p}{p\sigma_x^p}\left(\frac{\left\lvert\frac{\Delta}{T\sigma_x}\right\rvert^{q-p}}{1+\left\lvert\frac{\Delta}{T\sigma_x}\right\rvert^{q-p}}\right)$	$\frac{\lvert\Delta\rvert^{p-1}}{\sigma_x^p}\frac{\left\lvert\frac{\Delta}{T\sigma_x}\right\rvert^{q-p}\left(\frac{q}{p}+\left\lvert\frac{\Delta}{T\sigma_x}\right\rvert^{q-p}\right)}{\left(1+\left\lvert\frac{\Delta}{T\sigma_x}\right\rvert^{q-p}\right)^2}\text{sign}(\Delta)$

Summary

- A convex potential: stable MAP estimate and a global minimum
- The GMRF blurs edges; convex alternatives preserve them better
- The QGGMRF $\rho(\Delta) = \frac{|\Delta|^p}{p}\left(\frac{|\Delta/T|^{q-p}}{1+|\Delta/T|^{q-p}}\right)$ grows like $|\Delta|^q$ below T and like $|\Delta|^p$ above it, with continuous first and second derivatives.
- The scale parameter σ_x matches the model to the range of the data. For the GGMRF, $\hat{\sigma}_x^p = \frac{1}{N}\sum_{\{s,r\}\in\mathcal{P}} b_{s,r}|x_s - x_r|^p$.

Reading: Chapter 6.3–6.4. Next lecture: MAP estimation with non-Gaussian priors.