Contents

Chapter 1

Continous Non-Gaussian MRF Models

- Selection of potential functions
  - Nonconvex potentials
  - Convex potentials
    - Properties of convex functions
    - Continuity of MAP estimator => convex
    - Lp norm
      - Scale invariant
      - Continuously differentiable
      - Strictly convex for $p \geq 1$.
    - Absolute values => L1 norm
      - Total variation
      - Tends to produce flat regions with sharp transitions
      - Optimization can be difficult.
    - Huber function and Generalize Huber function
      - Quadratic near zero; linear far from zero
      - Preserves low contrast detail
      - Allows sharp edges to form
- Parameter estimation for MRFs
  - Estimation of general scale parameters
  - Estimation of parameters for GGMRFs

- MAP estimation with Pair-wise Non-Gaussian prior
  - Properties of convex optimization
  - ICD updates and rooting functions

In the previous chapters, we have introduced the GMRF and shown how it can be used as an image model in applications such as image restoration or reconstruction. However, one major limitation of the GMRF is that it does not accurately model the edges and other discontinuities that often occur in real images. In order to overcome this limitation, we must extend our approach to non-Gaussian models because edges are fundamentally not Gaussian.

The purposed of this chapter is to generalize the GMRF of the previous chapters to non-Gaussian random fields. To do this, we will introduce the concept of a pair-wise Gibbs distribution, and show that, while it includes the GMRF, it can be used to naturally generalize the GMRF to non-Gaussian distributions.

1.1 Continuous MRFs Based on Pair-Wise Cliques

In natural images, neighboring pixels typically have similar values. In this section, we develop image models with explicitly represent the probably of a particular image in terms of the differences between neighboring pixels. In order to do this, we first must reformulate the GMRF model so that it is expressed in terms of pixel difference. Once this is done, we can then generalize the GMRF model to non-Gaussian distributions.

First, let us review the concepts of neighborhood system, ∂s, and its associated pair-wise cliques, $\mathcal{P}$, from Section **??**. Recall, that $\partial s \subset S$ is the set of neighboring pixels to s, with the property that $r \in \partial s$ if and only if $s \in \partial r$. For a given neighborhood system, $\mathcal{P}$ denotes the set of all unordered neighboring pixel pairs $\{s, r\}$ such that $r \in \partial s$.

Using these conventions, the distribution of a GMRF can be written as

$$p(x) = \frac{1}{z} \exp\left\{ -\frac{1}{2} x^t B x \right\} , \tag{1.1}$$

where $B_{r,s} = 0$ when $r \notin \partial s$, and z is the normalizing constant for the distribution known as a partition function.

We would like to rewrite this expression in a way that makes the dependence on differences between neighboring pixels more explicit. To do this, we introduce a very useful vector-matrix relationship. For any matrix B and vector x, the following identity holds

$$x^t B x = \sum_{s\in S} a_s x_s^2 + \frac{1}{2}\sum_{s\in S}\sum_{s\in S} b_{s,r}|x_s - x_r|^2 \ , \tag{1.2}$$

where $a_s = \sum_{s\in S} B_{s,r}$ and $b_{s,r} = -B_{s,r}$.

When used to model images, the coefficients, a_s are most often chosen to be zero. By choosing $a_s = 0$, we insure that changing the mean of an image will not change its prior probability. To see this, consider the image x and a mean shift image, $\tilde{x} = x + 1$. If $a_s = 0$, then can be easily shown that

$$\tilde{x}^t B \tilde{x} = x^t B x \ ,$$

so the two images are equally probable.[1] By dropping these terms, and using the identity of (1.2), we get the following **pair-wise GMRF** distribution.[2]

$$p(x) = \frac{1}{z}\exp\left\{ -\sum_{\{s,r\}\in\mathcal{P}} b_{s,r}|x_s - x_r|^2 \right\} \ , \tag{1.3}$$

Notice that the distribution of (1.3) is explicitly written in terms of the differences between neighboring pixels. So if $b_{s,r} > 0$, then as the squared pixel difference $|x_s - x_r|^2$ increases, the probability decreases. This is reasonable since we would expect nearby pixels to have similar values.

However, a limitation of the GMRF models is that it can excessively penalize the differences between neighboring pixels. This is because, in practice, the value of $|x_s - x_r|^2$ can become very large when x_s and x_r fall across a discontinuous boundary in an image.

A simple solution to this problem is to replace the function $|x_s - x_r|^2$ with a new function $\rho(x_s - x_r)$, which grows less rapidly. For this purpose, let Δ denote the pixel difference, $x_s - x_r$. Then $\rho(\Delta)$ can be any positive,

[1] However, it should be noted that in this case the determent of the inverse covariance matrix B becomes zero, which is technically not allowed. However, this technical problem can always be resolved by computing MAP estimates with $a_s > 0$ and then allowing a_s to go to zero.

[2] Again, we note the the value of the partition function, z, in this expression becomes infinite because of the singular eigenvalue in B. However, this technical problem can be resolved by taking the limit as $a_s \to 0$.

continuous function with the following three properties.

$$\begin{aligned} \text{Zero at origin:} \quad & \rho(0) = 0 \\ \text{Symmetric:} \quad & \rho(-\Delta) = \Delta \\ \text{Monotone increasing:} \quad & \rho(|\Delta| + \epsilon) \geq \rho(|\Delta|) \text{ for } \epsilon > 0 \end{aligned}$$

Using this approach, we can generalize the Gaussian distribution of (1.3) to form a new distribution which we will refer to as a **pair-wise Gibbs distribution** given by

$$p(x) = \frac{1}{z} \exp\left\{ - \sum_{\{s,r\} \in \mathcal{P}} b_{s,r} \rho\left(\frac{x_s - x_r}{\sigma_x} \right) \right\} , \tag{1.4}$$

where we have reparameterized the distribution so that $b_{s,r}$ represent linear weights, and σ_x is a scaling parameter that controls the range of variation in x. So if our objective is to model images with large variations in pixel values, then σ_x can be increase proportionately.

In the expression for the pair-wise Gibbs distribution, we refer to the function $\rho(\Delta)$ as a **potential function**, and we refer to the derivative of the $\rho'(\Delta) = \frac{d\rho(\Delta)}{d\Delta}$ as its associated **influence function**. We will later see that the choice of potential and influence function can greatly impact the result when a pair-wise Gibbs distribution is used as a prior model in applications such as MAP estimation.

Importantly, the new non-Gaussian pair-wise MRF of equation (1.4) retains the most important property of the original GMRF, that is each pixel, x_s, is conditionally independent of the remaining pixels given its neighbors. To see this, we can factor the pair-wise Gibbs distribution into two portions: the portion that depends on x_s, and the portion that does not.

$$\begin{aligned} p(x) &= \frac{1}{z} \exp\left\{ - \sum_{\{s,r\} \in \mathcal{P}} b_{s,r} \rho(x_s - x_r) \right\} \\ &= \frac{1}{z} \exp\left\{ - \sum_{r \in \partial s} b_{s,r} \rho(x_s - x_r) \right\} f(x_{r \neq s}) \end{aligned}$$

where $f(x_{r \neq s})$ is some function of the pixels x_r for $r \neq s$. Using the factored form, we may calculate the conditional distribution of X_s given the remaining pixels X_r for $r \neq s$ as

$$\begin{aligned} p_{x_s|x_{r\neq s}}(x_s|x_{r\neq s}) &= \frac{p(x)}{\int_{\Re} p(x_s, x_{r\neq s})dx_s} = \frac{p(x_s, x_{r\neq s})}{\int_{\Re} p(x_s, x_{r\neq s})dx_s} \\ &= \frac{\frac{1}{z}\exp\left\{-\sum_{r\in\partial s} b_{s,r}\rho(x_s - x_r)\right\} f(x_{r\neq s})}{\int_{\Re} \frac{1}{z}\exp\left\{-\sum_{r\in\partial s} b_{s,r}\rho(x_s - x_r)\right\} f(x_{r\neq s})dx_s} \\ &= \frac{\exp\left\{-\sum_{r\in\partial s} b_{s,r}\rho(x_s - x_r)\right\}}{\int_{\Re} \exp\left\{-\sum_{r\in\partial s} b_{s,r}\rho(x_s - x_r)\right\} dx_s} \ . \end{aligned}$$

However, since this result is only a function of x_s and its neighbors, we have that

$$p_{x_s|x_{r\neq s}}(x_s|x_{r\neq s}) = p_{x_s|x_{\partial s}}(x_s|x_{\partial s}) \ .$$

This type of local dependency in the pixels of X is very valuable. In fact, it is valuable enough that it is worth giving it a name.

Definition: A random field, X_s for $s \in S$ is said to be a **Markov random field (MRF)** if for all $s \in S$, the conditional distribution of X_s is only dependent on its neighbors, i.e.,

$$p_{x_s|x_{r\neq s}}(x_s|x_{r\neq s}) = p_{x_s|x_{\partial s}}(x_s|x_{\partial s}) \ ,$$

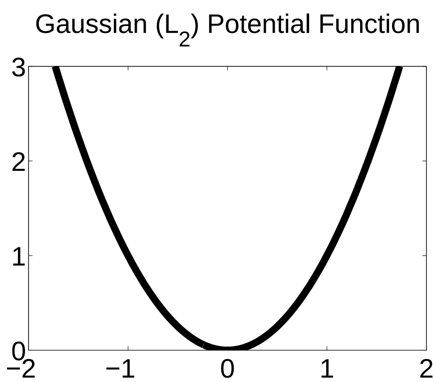
where $x_{\partial s}$ denotes the set of x_r such that $r \in \partial s$.

Using this definition, we may state the following very important property of pair-wise Gibbs distributions.

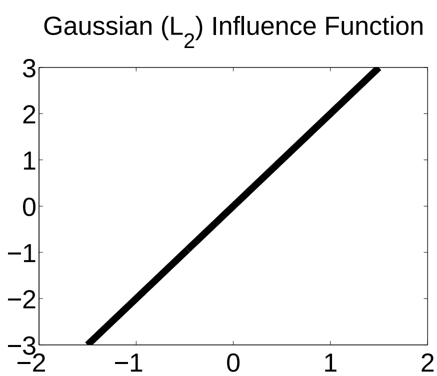
Property 1.1: *Pair-wise Gibbs distributions are the distributions of a MRF)* - Let X_s have a Gibbs distribution, $p(x)$, using neighborhood system ∂s and pair-wise cliques $\mathcal{P}$, then X_s is an MRF with neighborhood system ∂s, and the conditional probability of a pixel given its neighbors is given by

$$p_{x_s|x_{\partial s}}(x_s|x_{\partial s}) = \frac{1}{z}\exp\left\{-\sum_{r\in\partial s} b_{s,r}\rho(x_s - x_r)\right\} \ . \tag{1.5}$$

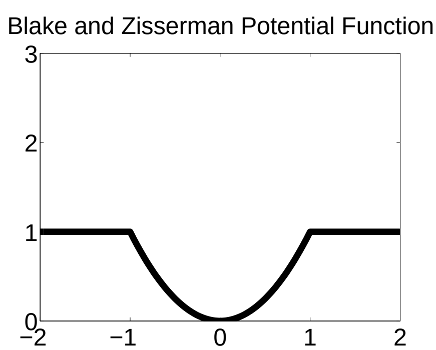
where $z = \int_{\Re} \exp\left\{-\sum_{r\in\partial s} b_{s,r}\rho(x_s - x_r)\right\} dx_s$.



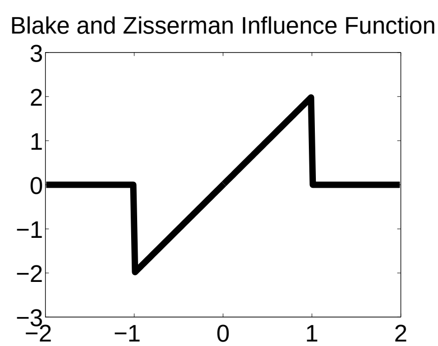
Gaussian (L_2) potential function



Gaussian (L_2) influence function



Weak spring potential function



Weak spring influence function

Figure 1.1: Figure showing both potential and influence functions for the Gaussian prior and weak-spring model. With the Gaussian prior, the influence of a pixel is unbounded, and it increases linearly with the value of Δ. However, with the weak spring model, the influence drops to zero when the value of $|\Delta|$ exceeds a threshold.

1.2 Selection of Potential and Influence Functions

The question remains of how to select a potential function that best models real images? One place to start is to look at the conditional distribution of a pixel, x_s, given its neighbors, $x_{\partial s}$. In particular, the most probable value of x_s given its neighbors is the solution to the optimization problem,

$$\begin{aligned} \hat{x}_s &= \arg\max_{x_s \in \Re} p_{x_s|x_{\partial s}}(x_s|x_{\partial s}) \\ &= \arg\min_{x_s \in \Re} \sum_{r \in \partial s} b_{s,r}\rho(x_s - x_r) \ , \end{aligned} \tag{1.6}$$

which can be in turn computed as the solution to the equation

$$\sum_{r \in \partial s} b_{s,r}\rho'(\hat{x}_s - x_r) = 0 \ . \tag{1.7}$$

Equation (1.6) has an interpretation of energy minimization where the terms $\rho(x_s - x_r)$ represent the potential energy associated with the two pixels x_s and x_r. Alternatively, equation (1.7) has the interpretation of a force balance equation where the terms $\rho'(\hat{x}_s - x_r)$ represent the force that the pixel x_r exerted on the estimate $\hat{x}_s$.

This interpretation of $\rho(x_s - x_r)$ as energy, and $\rho'(\hat{x}_s - x_r)$ as force serves as a useful tool for the design and selection of these potential functions. In fact, the function $\rho'(\Delta)$ is usually referred to as the **influence function** because it determines how much influence a pixel x_r has on the MAP estimate of a pixel $\hat{x}_s$.

Figure 1.1 shows the potential and influence functions for two possible choices of the potential function which we will refer to as the **Gaussian and Weak-spring potential [3, 2] functions**.

$$\begin{aligned} \text{Gaussian potential:} \quad & \rho(\Delta) = |\Delta|^2 \\ \text{Weak-spring potential:} \quad & \rho(\Delta) = \min\left\{|\Delta|^2, 1\right\} \end{aligned}$$

Along with each potential function, we also show the associated influence function, since this gives an indication of the "force" resulting from neighboring pixels with different values. Notice that with the Gaussian prior, the influence of a pixel is linearly proportional to the value of $\Delta = x_s - x_r$. Therefore, the influence of x_r on the estimate of $\hat{x}_s$ is unbounded, and a neighboring pixel on the other side of an edge can have and unbounded effect on the final estimate of $\hat{x}_s$.

However, with the weak spring potential, the influence of a neighboring pixel is bounded because the potential function is clipped to the value 1. This bounded influence of neighboring pixels is clearly shown by the associated influence function, which goes to zero for $|\Delta| > 1$. This means that pixel differences greater than σ_x have no influence on the final estimate of $\hat{x}_s$. If the pixels lie across a boundary or edge in an image, this lack of influence can be very desirable, since it minimizes any blurring and therefore preserves the edge detail.

Figure 1.2 shows some additional potential functions with a similar nature to the weak-spring potential. For each potential function, there is a point beyond which the potential function stops increasing at the same rate as a quadratic, and at this point, the influence functions magnitude begins to decrease. While these potential functions can preserve edges better than a GMRF model, they also tend to generate discontinuous estimates at image boundaries which are typically perceived as undesirable visual artifacts. For some applications, such as edge detection, this type of abrupt change at edges can be desirable. However, in applications where the output needs to be viewed as an physical image (i.e. photography, medical imaging, etc.), then this type of very abrupt change can be very undesirable.

1.3 Convex Potential Functions

Appendix **??** presents some basic facts about convex functions and their use in optimization. Using the results of this appendix, we can see that the cost function to be optimized

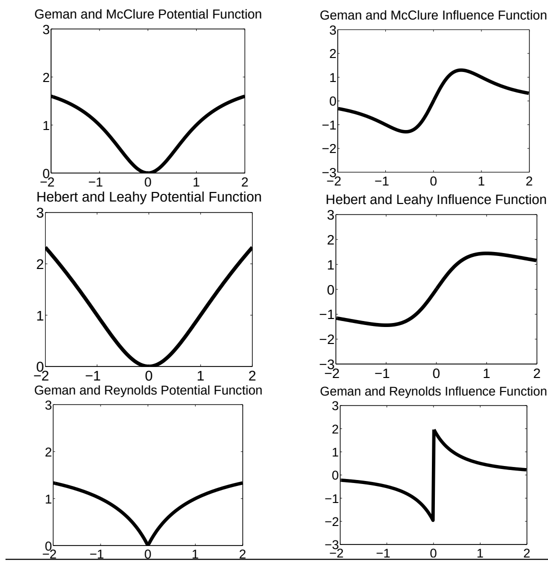
$$f(x) = ||y - Ax||^2 + \lambda x^t B x$$

is strictly convex and has a unique global minimum at $\mu(y)$ which is a solution to the equation

$$\nabla f(x)|_{x=\hat{x}} = 0 \ .$$

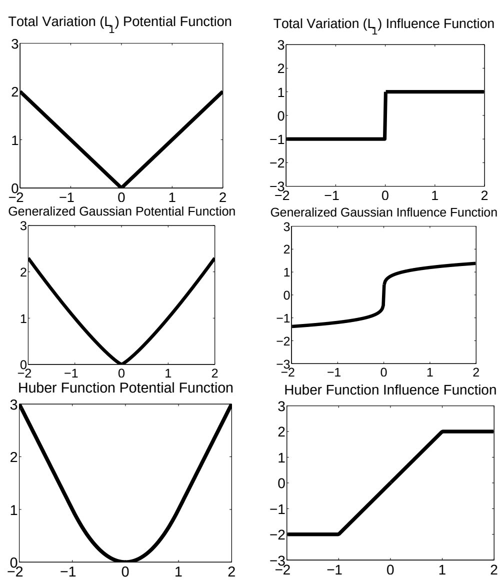
****

Continuous (Stable) MAP Estimation[4]



Potential Function	Influence Function	$\rho(\Delta)$	Reference
Geman and McClure	Geman and McClure	$\rho(\Delta) = \frac{\Delta^2}{1+\Delta^2}$	[7, 8]
Hebert and Leahy	Hebert and Leahy	$\rho(\Delta) = \log\left(1+\Delta^2\right)$	[9]
Geman and Reynolds	Geman and Reynolds	$\rho(\Delta) = \frac{\lvert\Delta\rvert}{1+\lvert\Delta\rvert}$	[6]

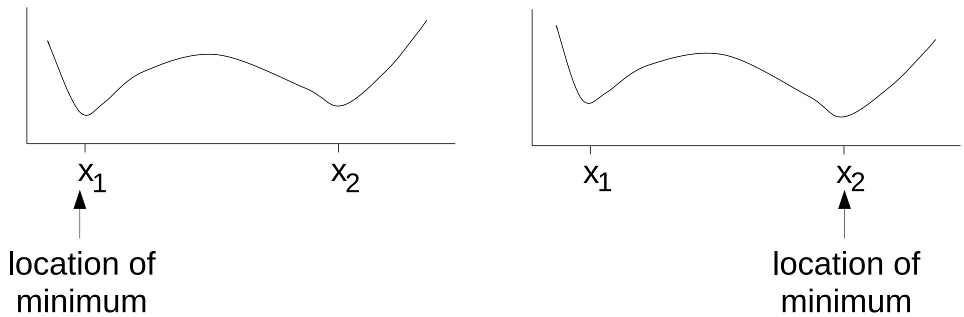
Figure 1.2: Figure showing both potential and influence functions for a variety of non-convex potential functions.



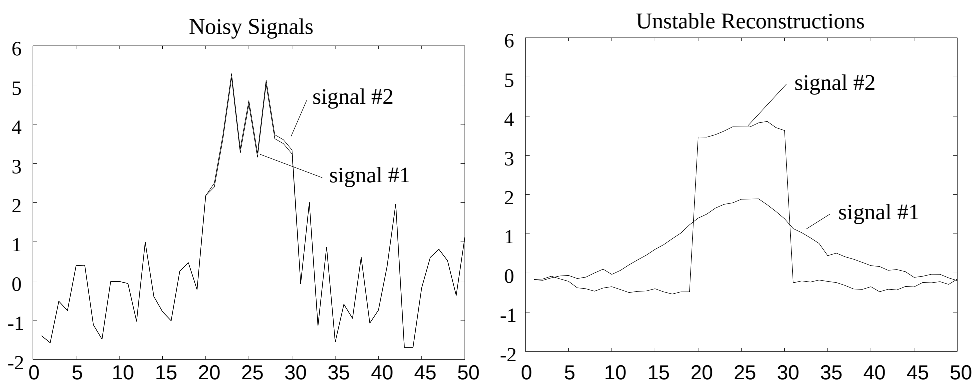
Potential Function	Influence Function	$\rho(\Delta)$	Reference
Total Variation (L_1)	Total Variation (L_1)	$\rho(\Delta) = \lvert\Delta\rvert$	[1]
Generalized Gaussian	Generalized Gaussian	$\rho(\Delta) = \lvert\Delta\rvert^p$	[4]
Huber Function	Huber Function	$\rho(\Delta) = \min\left\{\lvert\Delta\rvert^2, 2\lvert\Delta\rvert - 1\right\}$	[11]

Figure 1.3: Figure showing both potential and influence functions for a variety of convex potential functions.

- Minimum of non-convex function can change abruptly.



- Discontinuous MAP estimate for Blake and Zisserman potential.



- Theorem:[4] - If the log of the posterior density is **strictly convex**, then the MAP estimate is a continuous function of the data.

Properties of Convex Potential Functions

- Both $\log\cosh(\Delta)$ and Huber functions
  - Quadratic for $|\Delta| << 1$
  - Linear for $|\Delta| >> 1$
  - Transition from quadratic to linear determines edge threshold.
- Generalized Gaussian MRF (GGMRF) functions
  - Include $|\Delta|$ function
  - Do not require an edge threshold parameter.
  - Convex and differentable for $p > 1$.

1.4 Parameter Estimation for Continuous MRFs

Parameter Estimation for Continuous MRF's

- Topics to be covered:
  - Estimation of scale parameter, σ
  - Estimation of temperature, T, and shape, p

ML Estimation of Scale Parameter, σ, for Continuous MRF's [5]

- For any continuous state Gibbs distribution

$$p(x) = \frac{1}{Z(\sigma)} \exp\left\{-U(x/\sigma)\right\}$$

the partition function has the form

$$Z(\sigma) = \sigma^N Z(1)$$

- Using this result the ML estimate of σ is given by

$$\left.\frac{\sigma}{N}\frac{d}{d\sigma}U(x/\sigma)\right|_{\sigma=\hat{\sigma}} - 1 = 0$$

- This equation can be solved numerically using any root finding method.

ML Estimation of σ for GGMRF's [10, 5]

- For a Generalized Gaussian MRF (GGMRF)

$$p(x) = \frac{1}{\sigma^N Z(1)} \exp\left\{-\frac{1}{p\sigma^p}U(x)\right\}$$

where the energy function has the property that for all $\alpha > 0$

$$U(\alpha x) = \alpha^p U(x)$$

- Then the ML estimate of σ is

$$\hat{\sigma} = \left(\frac{1}{N}U(x)\right)^{(1/p)}$$

- Notice for that for the i.i.d. Gaussian case, this is

$$\hat{\sigma} = \sqrt{\frac{1}{N}\sum_s |x_s|^2}$$

Chapter 1 Problems

1. Let $f : I\!R^N \rightarrow I\!R$ and $g : I\!R^N \rightarrow I\!R$ be convex functions, and let A be an $N \times M$ matrix. Prove that
   a) $h(x) = f(Ax)$ for $x \in I\!R^M$ is a convex function.
   b) $h(x) = f(x) + g(x)$ is a convex function.
   c) $f(x)$ is continuous.

2. Let $f : I\!R^N \rightarrow I\!R$ be a strictly convex continuously differentiable function. Prove that if $\bigtriangledown f(x^*) = 0$ if and only if x^* is the unique global minimum of $f(\cdot)$.

3. Find a convex function $f(x_1, x_2)$ with a unique global minimum, so that coordinate decent does not converge to the global minimum. Why?

4. Let $f : I\!R^N \rightarrow I\!R$ and $g : I\!R^N \rightarrow I\!R$ be convex functions, let A be an $N \times M$ matrix, and let B be an $N \times N$ symmetric positive definite matrix. Prove that
   a) $h(x) = f(Ax)$ for $x \in I\!R^M$ is a convex function.
   b) $h(x) = f(x) + g(x)$ is a convex function.
   c) $f(x)$ is continuous.
   d) $f(x) = ||x||^2$ is strictly convex.
   e) $f(x) = ||x||_B^2 = x^t B x$ is strictly convex.

5. Let $f : I\!R^N \rightarrow I\!R$ be a strictly convex continuously differentiable function. Prove that if $\bigtriangledown f(x^*) = 0$ if and only if x^* is the unique global minimum of $f(\cdot)$.

6. Consider the optimization problem

$$\hat{x} = \arg \min_{x \in \Re^N} \left\{ ||y - Ax||^2 + \lambda x^t B x \right\}$$

   where A is a nonsingular $N \times N$ matrix, B is a positive definite $N \times N$ matrix, and $\lambda > 0$.
   a) Show that the cost function is strictly convex.
   b) Derive a closed form expression for the solution.

7. Show that the conditional probabilities of equation (?) and equation (**??**) are equivalent.

8. Let B be any $N \times N$ symmetric matrix, and let x by an N dimensional vector. Let $\mathcal{P}$ be the set of all (unique) sets of pixel pairs

$$\mathcal{P} = \{\{i,j\} : \text{for } 0 \le i,j \le N\ \} \ ,$$

and define $a_i = \sum_{j=1}^{N} B_{i,j}$ and $b_{i,j} = -B_{i,j}$.

a) Show that the following equality holds.

$$x^t B x = \sum_{i=1}^{N} a_i x_i^2 + \sum_{\{i,j\}\in\mathcal{P}} b_{i,j}(x_i - x_j)^2$$

b) Use this to show the result of Property 1.1.

9. Show that the costs resulting from ICD updates forms a monotone decrease sequence that is bounded below.

10. Show that any local minimum of the cost function is also a global minimum.

11. Use the monochrome image img04.tif as x and produce a noisy image y by adding i.i.d. Gaussian noise with mean zero and $\sigma_W^2 = 16^2$. Approximate y by truncating the pixels to the range $[0, \cdots 255]$. Print out the image Y.

12. Compute the MAP estimate of X using 20 iterations of ICD optimization. Use $\sigma_x^2 = \hat{\sigma}_x^2$ the ML estimate of the scale parameter computed for $p = 2$, and $\sigma_W^2 = 16^2$. Print out the resulting MAP estimate.

13. Plot the cost function as a function of the iteration number for the experiment of step 12.

14. Repeat step 12 for $\sigma_x^2 = 5 * \hat{\sigma}_x^2$, and $\sigma_x^2 = (1/5) * \hat{\sigma}_x^2$.

15. Compute the noncausal prediction error for the image img04.tif

$$e_i = x_i - \sum_{j\in\partial i} g_{i-j} x_j$$

and display it as an image by adding an offset of 127 to each pixel. Clip any value which is less than 0 or greater than 255 after adding the offset of 127.

16. Compute $\hat{\sigma}_{ML}$ the ML estimate of the scale parameter σ for values of p in the range $0.1 \leq p \leq 2$. Do not include cliques that fall across boundaries of the image. Plot $\hat{\sigma}_{ML}$ (not $\hat{\sigma}^p_{ML}$) versus p for p ranging from 0.1 to 2.

Bibliography

[1] J. Besag. Towards Bayesian image analysis. *Journal of Applied Statistics*, 16(3):395–407, 1989.

[2] A. Blake. Comparison of the efficiency of deterministic and stochastic algorithms for visual reconstruction. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 11(1):2–30, January 1989.

[3] A. Blake and A. Zisserman. *Visual Reconstruction*. MIT Press, Cambridge, Massachusetts, 1987.

[4] C. A. Bouman and K. Sauer. A generalized Gaussian image model for edge-preserving MAP estimation. *IEEE Trans. on Image Processing*, 2(3):296–310, July 1993.

[5] C. A. Bouman and K. Sauer. Maximum likelihood scale estimation for a class of Markov random fields. In *Proc. of IEEE Int'l Conf. on Acoust., Speech and Sig. Proc.*, volume 5, pages 537–540, Adelaide, South Australia, April 19-22 1994.

[6] D. Geman and G. Reynolds. Constrained restoration and the recovery of discontinuities. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 14(3):367–383, March 1992.

[7] S. Geman and D. McClure. Bayesian images analysis: An application to single photon emission tomography. In *Proc. Statist. Comput. sect. Amer. Stat. Assoc.*, pages 12–18, Washington, DC, 1985.

[8] S. Geman and D. McClure. Statistical methods for tomographic image reconstruction. *Bull. Int. Stat. Inst.*, LII-4:5–21, 1987.

[9] T. Hebert and R. Leahy. A generalized EM algorithm for 3-D Bayesian reconstruction from Poisson data using Gibbs priors. *IEEE Trans. on Medical Imaging*, 8(2):194–202, June 1989.

[10] K. Lange. An overview of Bayesian methods in image reconstruction. In *Proc. of the SPIE Conference on Digital Image Synthesis and Inverse Optics*, volume SPIE-1351, pages 270–287, San Diego, CA, 1990.

[11] R. Stevenson and E. Delp. Fitting curves with discontinuities. *Proc. of the first international workshop on robust computer vision*, pages 127–136, October 1-3 1990.