# ON PARAMETER ESTIMATION FOR A CLASS OF MARKOV RANDOM FIELD IMAGE MODELS:

Sean Borman

April 12, 1996

## Abstract

Mathematical image models find common application as *a-priori* information in many image reconstruction and restoration problems. A broad class of models, Markov random fields, are well recognized for their convenient properties in these problems. The characteristic behavior of a MRF is determined by its set of model *parameters*. Appropriate choice or *estimation* of these parameters is necessary to ensure that the model is a faithful representation of the *a-priori* knowledge.

*Estimation* of the *clique weight parameters* of a particular class of MRF, the generalized Gaussian MRF is the primary topic investigated in this thesis. In the proposed approach, minimization of prediction error estimation (MPEE), we choose the clique weight parameters for which the penalized error between model predicted data and known realization data is a minimum. Procedures for minimization of this error are proposed and the peformance of the MPEE method is evaluated.

# INTRODUCTION

Primary Research Problem...

*ESTIMATION* of the *CLIQUE WEIGHT PARAMETERS* for the generalized Gaussian Markov random field (GGMRF)

STEPS TO GET THERE...

- Background
  - preliminary knowledge required understand problem.
- Simulation
  - simulating MRF's to obtain realizations
- Estimation
  - the "good stuff"
  - estimating clique weights from realization data
- Conclusions

# BACKGROUND - Image Models

- What's an image model?
  - "capture" behavior of required image class.
- Why do we want one anyway ?
  - short answer ... they're useful!
- Applications
  - too many specific applications!
  - focus on one useful *framework*
  - stochastic regularization and Bayesian estimation

# BACKGROUND - Bayesian estimation 1

- Observation model

$$\mathbf{Y} = \mathbf{H}\mathbf{Z} + \mathbf{N}, \tag{1}$$

  - $\mathbf{Y}$ - observed output
  - $\mathbf{H}$ - known linear transformation
  - $\mathbf{Z}$ - unknown data
  - $\mathbf{N}$ - additive noise, known distribution

- Seek $\hat{\mathbf{Z}}$ - estimate of $\mathbf{Z}$

- Direct approach ?

$$\begin{aligned} \hat{\mathbf{Z}} &= \mathbf{H}^{-1}\mathbf{Y} \\ \rightarrow \hat{\mathbf{Z}} &= \mathbf{Z} + \mathbf{H}^{-1}\mathbf{N}. \end{aligned} \tag{2}$$

  - maybe, if everything is well behaved
  - what if $\mathbf{H}$ singular?
  - or $\mathbf{H}$ is poorly conditioned?
  - what if there is insufficient observation data $\mathbf{Y}$?
  - $\rightarrow$ ill-posed problem

# BACKGROUND - Bayesian estimation 2

- Hadamard's definition of a "well-posed" problem
  1. solution exists
  2. is unique
  3. depends continuously on the data

  Often (2) is violated.

- What can we do to fix the problem ?
  - regularized solution methods
  - incorporate *a-priori* information about $\mathbf{Z}$
  - ensures Hadamard constraints are met
  - deterministic methods
  - stochastic methods
    * effected using Bayesian framework
    * MAP estimate of $\mathbf{Z}$

# BACKGROUND - Bayesian estimation 3

- Bayesian approach for estimation of $\mathbf{Z}$
  - maximization of *a-posteriori* probability $\Pr(\mathbf{Z}|\mathbf{Y})$

$$\hat{\mathbf{Z}}_{MAP} = \arg\max_{\mathbf{Z}} \left\{ \Pr(\mathbf{Z}|\mathbf{Y}) \right\}. \quad (3)$$

  - apply Bayes' rule

$$\hat{\mathbf{Z}}_{MAP} = \arg\max_{\mathbf{Z}} \left\{ \frac{\Pr(\mathbf{Y}|\mathbf{Z}) \Pr(\mathbf{Z})}{\Pr(\mathbf{Y})} \right\}. \quad (4)$$

  - $\hat{\mathbf{Z}}_{MAP}$, is unaffected by constant term $\Pr(\mathbf{Y})$

$$\hat{\mathbf{Z}}_{MAP} = \arg\max_{\mathbf{Z}} \left\{ \Pr(\mathbf{Y}|\mathbf{Z}) \Pr(\mathbf{Z}) \right\}. \quad (5)$$

  - log is monotonic, increasing, so equivalently solve

$$\hat{\mathbf{Z}}_{MAP} = \arg\max_{\mathbf{Z}} \left\{ \log \Pr(\mathbf{Y}|\mathbf{Z}) + \log \Pr(\mathbf{Z}) \right\}. \quad (6)$$

# BACKGROUND - Bayesian estimation 4

- Two terms in the conditional probability

  1. $\Pr(\mathbf{Y}|\mathbf{Z})$ – familiar in ML estimation
  2. $\Pr(\mathbf{Z})$ – the *a-priori probability* or *prior* provides a "bias" term which favors solutions consistent with the model of the underlying process $\mathbf{Z}$

- Inclusion of the additional *a-priori* information $\Pr(\mathbf{Z})$ makes for a "regularized" solution

- Estimate $\hat{\mathbf{Z}}_{MAP}$ balances fidelity to the observed data as well as the *a-priori* model.

# BACKGROUND - MRF Nomenclature 1

- Random fields on finite point sets
  - set of $M$ "sites" $\mathcal{S} \triangleq \{S_i : 1 \leq i \leq M\}$.
  - example: lexicographic ordering of pixel sites
  - $\forall\ S_i \in \mathcal{S},\ \exists$ random variable $Z_i$
  - define random vector $\mathbf{Z}$ from ordered set of $Z_i$
  - determine sample space $\Omega$ based on $\mathcal{S}$
  - determine Borel set $\Psi$ on $\Omega$
  - specify $\Pr$, a probability measure on $\Psi$
  - $\rightarrow$ valid *probability space*

- Neighbors
  - define $\forall\ S_i \in \mathcal{S}$, a set of *neighbors* $\mathcal{N}_i$
  - example: geometric 4,8,etc neighbors
  - a site cannot be its own neighbor
  - $S_j \in \mathcal{N}_i \rightarrow S_i \in \mathcal{N}_j$
  - *neighborhood system* $\mathcal{N} \triangleq \{\mathcal{N}_i : 1 \leq i \leq M\}$

# BACKGROUND - MRF Nomenclature 2

- Model Order

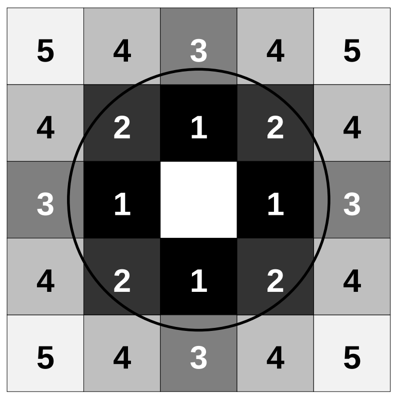



Figure 1: Neighbors of given order are specified as the set of pixels either partially or fully contained by circle of increasing radius centered on a given pixel.

# BACKGROUND - MRF Nomenclature 3

- Cliques
  - a clique is a set of sites, all pairs of which are mutual neighbors.
  - denote the clique of $S_i$ as $\mathcal{C}_i$
  - the set of all cliques is denoted $\mathcal{C}$

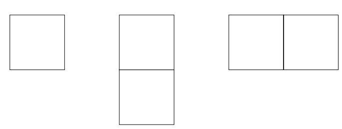


Figure 2: Cliques of a first order neighborhood.

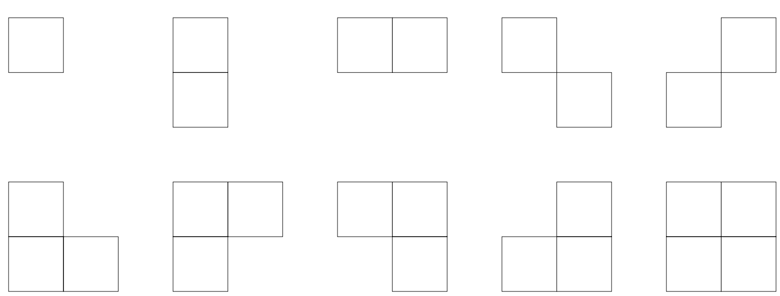


Figure 3: Cliques of a second order neighborhood.

# BACKGROUND - MRF Definition

- $\mathbf{Z}$, the ordered collection of random variables of site values, is a Markov random field with respect to its neighborhood system provided that for every $S_i \in \mathcal{S}$ and realization $\omega$ in the sample space $\Omega$,

$$\Pr(\mathbf{Z} = \omega) \; > \; 0 \;\; \forall \, \omega \in \Omega \tag{7}$$

$$\begin{aligned} \Pr(Z_i = z_i \mid Z_k = z_k, \; k \neq i) \;\; &= \\ \Pr(Z_i = z_i \mid Z_k = z_k, \; k \in \mathcal{N}_i). \end{aligned} \tag{8}$$

- Second equation lends the name “Markov”

- Question: what is the form of the joint PDF of $\mathbf{Z}$ ?

# BACKGROUND - MRF/Gibbs Equivalence 1

- A MRF has a Gibbs distribution
  - random vector $\mathbf{Z}$ associated with sites $\mathcal{S}$ and neighborhood system $\mathcal{N}$ is a MRF iff

$$\Pr(\mathbf{Z}) = \frac{1}{K_p} \exp \left\{ -\frac{1}{\beta} \sum_{c \in \mathcal{C}} V_c(\mathbf{Z}) \right\}. \quad (9)$$

  - this is known as the Hammersley Clifford theorem

# BACKGROUND - MRF/Gibbs Equivalence 2

- $K_p$ - the *partition function*
  - $K_p$ is constant
  - found by integrating the un-scaled PDF expression over all possible states of the argument $\mathbf{Z}$
  - very difficult to compute
  - complicates estimation problems involving MRF's.

- $\beta$ - the *"temperature"* parameter
  - akin to scaled *temperature* parameter $k_B T$ of the Gibbs distribution encountered in thermodynamics.

- $V_c(\cdot)$ - the *clique function*
  - function of the MRF cliques
  - $\mathcal{C}$ denotes the set of all image cliques
  - summation of clique function terms gives the *energy* of the field.

# BACKGROUND - More Nomenclature

- Refine definition
  - $\mathcal{A}_c$ - *clique activity function*
  - $\rho_{c,\alpha}(\cdot)$ - *clique activity penalty function*
  - $w_c$ - *clique weight*

$$\Pr(\mathbf{Z}) = \frac{1}{K_p} \exp\left\{ -\frac{1}{\beta} \sum_{c \in \mathcal{C}} w_c \cdot \rho_{c,\alpha}\left(\mathcal{A}_c(\mathbf{Z})\right) \right\}. \tag{10}$$

- Further refinements
  - $w_c$ often constant for each clique type
  - homogeneous field - single clique activity penalty function for each clique type
  - pairwise interaction MRF - clique activity penalty function same for all cliques
  - $i$ numbers each unique type of clique pair in the set of pair cliques $\mathcal{C}_2$.

$$\Pr(\mathbf{Z}) = \frac{1}{K_p} \exp\left\{ -\frac{1}{\beta} \sum_{c \in \mathcal{C}_2} w_i \cdot \rho_{\alpha}\left(\mathcal{A}_c(\mathbf{Z})\right) \right\}. \tag{11}$$

# BACKGROUND - Common MRF's 1

- Gaussian MRF
  - excellent representation of smooth data
  - poor edge representation

$$\rho(x) = x^2. \tag{12}$$

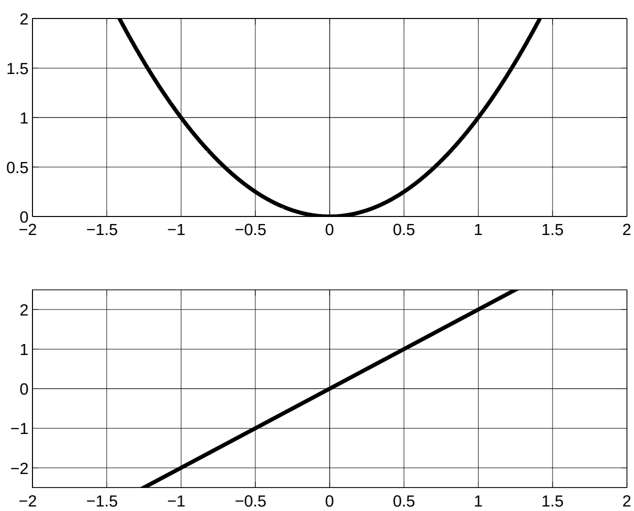



Figure 4: Gaussian MRF clique activity penalty function and derivative.

- Discontinuity preserving MRF's

# BACKGROUND - Common MRF's 2

- Blake and Zisserman
  - excellent edge representation
  - good smoothing
  - non-convex $\rightarrow$ difficulties in optimizations

$$\rho_\alpha(x) = \begin{cases} x^2, & |x| \leq \alpha, \\ \alpha^2, & |x| > \alpha. \end{cases} \quad (13)$$

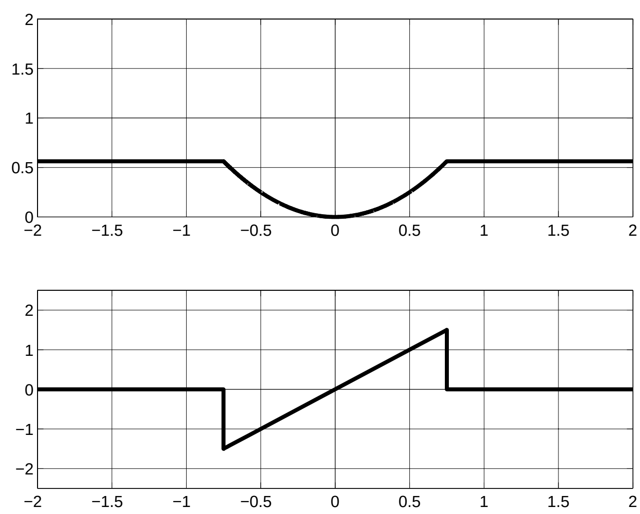



Figure 5: Blake and Zisserman type clique activity penalty function and derivative.

# BACKGROUND - Common MRF's 3

- Huber MRF
  - good edge representation
  - good smoothing
  - convex

$$\rho_\alpha(x) = \begin{cases} x^2, & |x| \le \alpha, \\ \alpha^2 + 2\alpha(|x| - \alpha), & |x| > \alpha. \end{cases} \qquad (14)$$

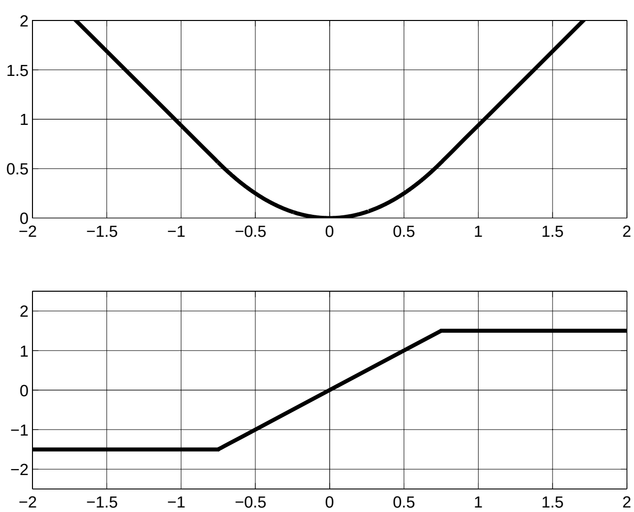



Figure 6: Huber type clique activity penalty function and derivative.

# BACKGROUND - Common MRF's 4

- Absolute Value MRF (AVMRF)
  - superior edge representation
  - inferior smoothing
  - convex

$$\rho(x) = |x|. \tag{15}$$

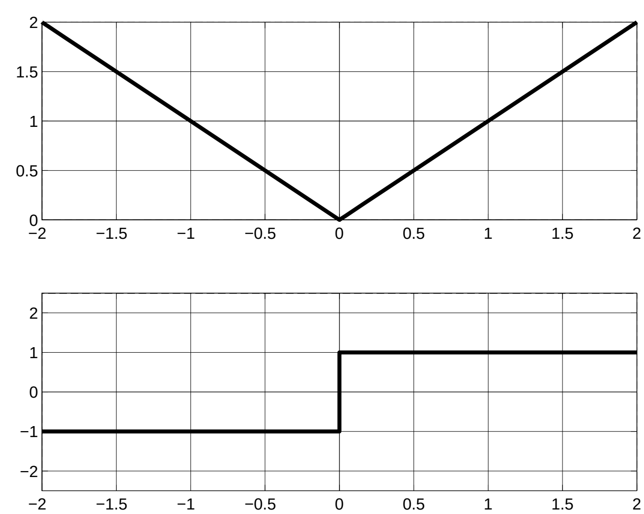



Figure 7: Absolute value clique activity penalty function and derivative.

# BACKGROUND - Common MRF's 5

- Generalized Gaussian MRF (GGMRF)
  - good edge representation
  - good smoothing
  - convex

$$\rho_\alpha(x) = |x|^\alpha, \qquad 1 \leq \alpha \leq 2, \tag{16}$$

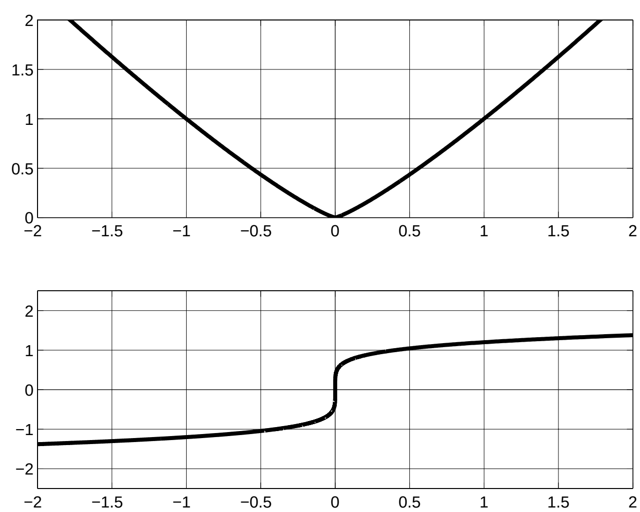



Figure 8: Generalized Gaussian MRF clique activity penalty function and derivative.

# SIMULATION - Why Simulation?

- Simple example - computing the mean $\mu_{\mathbf{z}}$ of a MRF
  - direct approach

$$\mu_{\mathbf{z}} = \int_{\Omega} \mathbf{z} \cdot \Pr(\mathbf{z}) \; d\mathbf{z}. \tag{17}$$

  - dimensionality of the sample space $\Omega$ ?
  - infeasible to apply direct methods
  - what if we do not have complete knowlege of the MRF parameters?
  - especially the partition function!

- An alternative strategy - simulation
  - can we simulate the PDF?
  - if so, we can find statistics by assuming ergodicity
  - replace ensemble averages by time averages
  - simulation provides *realizations* from the MRF
  - thse realizations may be used for testing purposes

# SIMULATION - Methods

- Metropolis algorithm
  - idea: construct Markov chain in time which converges to the equilibrium distribution of the desired MRF PDF.
  - desired PDF is $\pi$
  - construct chain with $\pi$ as equilibrium distribution
  - distribution of the chain after $n$ simulation iterations is $\pi_n$
  - then $\pi_n \to \pi$ as $n$ increases.

- How do we implement it?
  1. choose an initial state $S_0$. e.g. assign random values to image pixels.
  2. given any state $S_i$, propose new state $S_{i+1}$. e.g. random increment added to some pixel.
  3. accept state $S_{i+1}$ with probability $\min\left(1, \frac{\Pr(S_{i+1})}{\Pr(S_i)}\right)$, else remain in state $S_i$.
  4. return to step 2, for some other pixel.

- Do not need to know partition function!

# SIMULATION - Clique Weight Conventions

- Convention for writing clique weights as vector
  - homogeneity - number of cliques is number of clique tupes
  - pairwise interaction MRF - further reduction
  - homogeneity implies symmetry of clique weights
  - introduce clique weight numbering convention
  - $2^{\text{nd}}$ order MRF: $[w_1\ w_2\ w_3\ w_4\ w_4\ w_3\ w_2\ w_1]^T$
  - use only unique part of vector $[w_1\ w_2\ w_3\ w_4]^T$

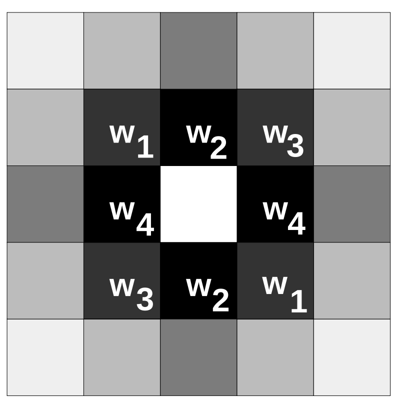



Figure 9: clique weight vector for $2^{\text{nd}}$ order MRF.

  - general neighborhood - raster scan from upper-most, left-most site.

# SIMULATION - Convergence 1

- How long is long enough?
  - Ripley and Kirkland
  - estimate suitable parameter from chain
  - monitor convergence behavior of estimate
  - convergence of estimate
    $\rightarrow$ convergence of chain

- Concentrate on the Generalized Gaussian MRF
  - Bouman and Sauer
  - $\hat{\sigma}$, MLE of GGMRF $\sigma$, may be found given $p$
  - $\sigma$ akin to $\beta$ - so called "scale" parameter
  - method:
    * begin simulation with known equilibrium $p, \sigma$
    * estimate $\hat{\sigma}$, given $p$ from successive realizations
    * terminate simulation on convergence of $\hat{\sigma}$ to $\sigma$.

# SIMULATION - Convergence 2

- Example of convergence of $\hat{\sigma}$

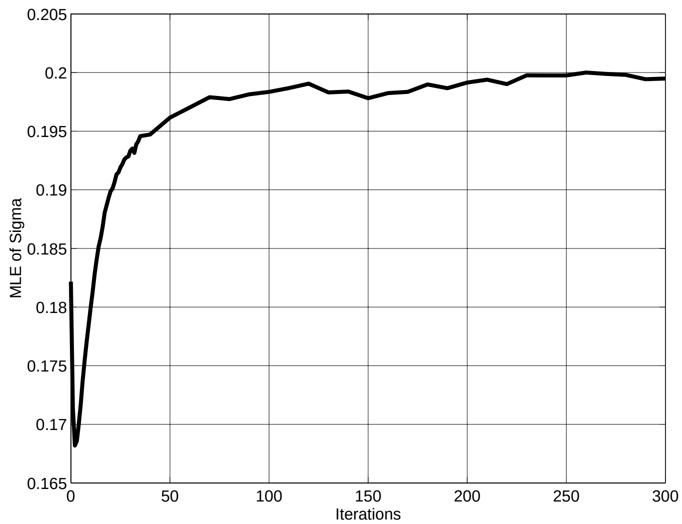



Figure 10: Convergence of MLE $\hat{\sigma}$ given $p$ for the simulation of realizations of a GGMRF with $\sigma = 0.2$ and $p = 1.1$, with clique weight vector $\frac{1}{8} \cdot [1\ 3]^T$.

# SIMULATION - Examples 1

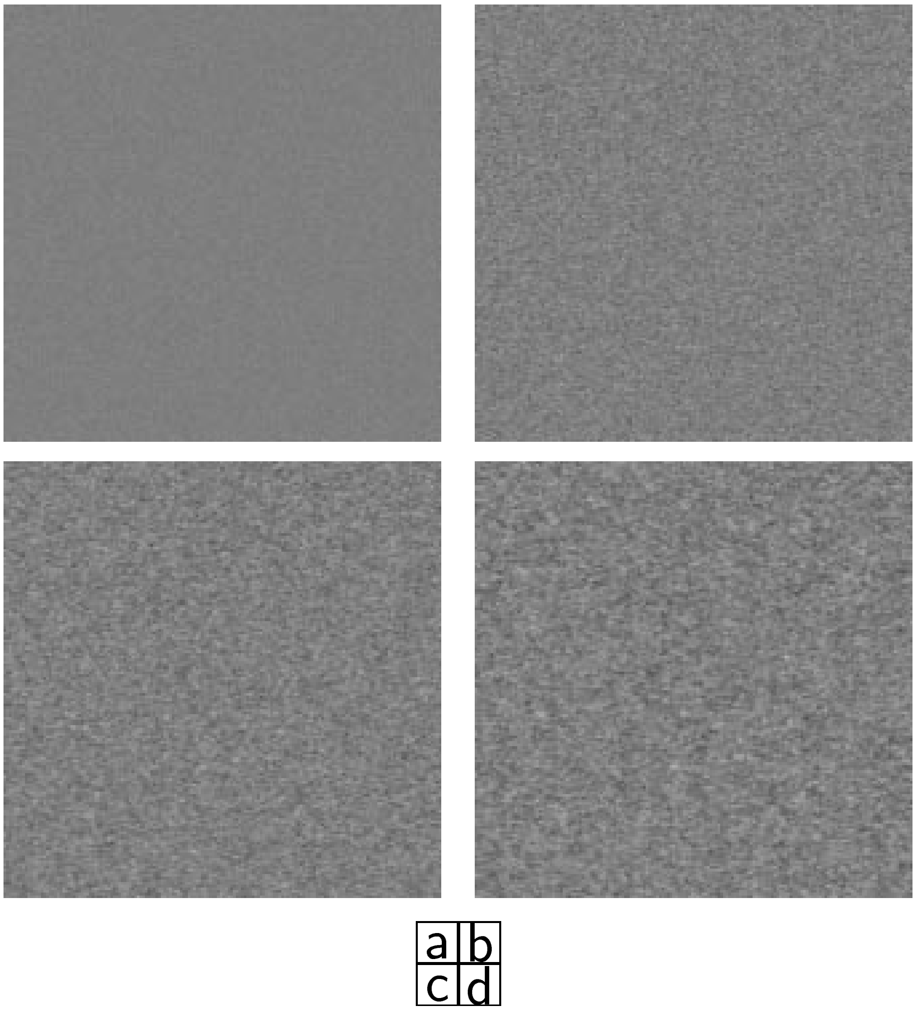
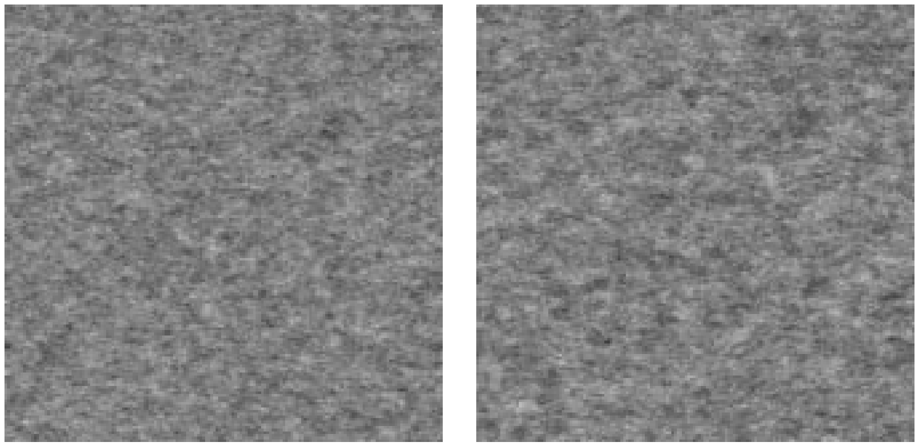



Figure 11: Sequence of images in generation of $1^{st}$ order GGMRF realization with $\sigma = 0.2$, $p = 1.1$ and clique weight vector $\frac{1}{8} \cdot [1\ 3]^T$. (a) Initial guess. (b) 20 iterations. (c) 50 iterations. (d) 100 iterations.

# SIMULATION - Examples 2



a | b

Figure 12: Sequence of images in generation of $1^{st}$ order GGMRF realization with $\sigma = 0.2$, $p = 1.1$ and clique weight vector $\frac{1}{8} \cdot [1\ 3]^T$. (a) 400 iterations. (b) 800 iterations.

# ESTIMATION

- Introduction
  - estimation of MRF parameters
  - why is this problem non-trivial?
  - existing approaches to MRF parameter estimation

- Our work
  - AVRMF
    * AVMRF and Weighted Order Statistic filters
    * clique weight parameter estimation by "Minimization of Prediction Error" (MPE)
    * MPE for the AVMRF
  - GGMRF
    * GGMRF and the "Generalized" WOS filter
    * MPE for the GGMRF
  - MPEE for other classes of MRF

# ESTIMATION - Introduction

- MRF's are parameterized
  - control the behavior of the field
- A common problem
  - given a realization with unknown parameters
  - how do we find the parameters from realization data?
- Estimation
  - our focus - clique weight parameters
  - find the parameters which "best fit" the realization

# ESTIMATION - Difficulties 1

- So what's the problem?
  - MRF is an exponential distribution
  - parameter estimation is therefore well known
  - so what's the problem?
  - Pickard:
    "*Likelihood inference is, therefore, apparently straightforward. The requisite normalizing constants, however, are obstreperous*."
  - partition function $K_p$

# ESTIMATION - Difficulties 2

- An example - the direct approach
  - ML estimation of weight parameter $w_1$
  - recall form of MRF PDF

$$\Pr(\mathbf{Z}) = \frac{1}{K_p} \exp\left\{ -\frac{1}{\beta} \sum_{c \in \mathcal{C}_2} w_i \cdot \rho_\alpha\left(\mathcal{A}_c(\mathbf{Z})\right) \right\}. \tag{18}$$

  - We set up the ML approach as

$$\hat{w}_1 = \arg\max_{w_1} \left\{ \Pr(\mathbf{Z}|w_1) \right\}, \tag{19}$$

  - which may be reduced to finding,

$$\hat{w}_1 = \arg\max_{w_1} \left\{ \ln\left(\frac{1}{K_p}\right) - \frac{1}{\beta} \sum_{c \in \mathcal{C}_2} w_1 \cdot \rho_\alpha\left(\mathcal{A}_c(\mathbf{Z})\right) \right\}. \tag{20}$$

  - $K_p$ is an integral over all states of the field and is thus also a function of $w_1$!
  - often $K_p$ is unknown and is infeasible to compute

# ESTIMATION - AVMRF & WOS Filters

- Relationship between AVMRF and WOS Filters
- Form of AVMRF

$$\Pr(\mathbf{Z}) = \frac{1}{K_p} \exp\left[ -\frac{1}{\beta} \sum_{c \in \mathcal{C}_2} w_i \cdot \rho_\alpha \left( \mathcal{A}_c(\mathbf{Z}) \right) \right] \quad (21)$$

  - $\rho_\alpha(\cdot) = |\cdot|$, $i$ numbers unique types of pair cliques
  - $\mathcal{C}_2$ is the set of pair cliques

- MAP estimate of pixel given neighbors
  - *maximize* $\Pr(Z_k|\mathbf{Z} \setminus Z_k)$
  - apply Markov property of PDF
  - $\Pr(Z_k|\mathbf{Z} \setminus Z_k) = \Pr(Z_k|Z_j)$ for $j \in \mathcal{N}_k$

$$\Pr(Z_k|\mathbf{Z} \backslash Z_k) = \frac{1}{{K_p}'} \exp\left[ -\frac{1}{\beta} \sum_{j \in \mathcal{N}_k} w_i |Z_k - Z_j| \right] \quad (22)$$

  - ${K_p}'$ is a new normalization constant

# ESTIMATION - AVMRF & WOS Filt. 2

- MAP estimate of pixel given neighbors ctnd . . .
  - find MAP estimate $\hat{Z}_k$, of $k^{\text{th}}$ pixel's value

$$\begin{aligned}
\hat{Z}_k &= \arg\max_{\zeta} \left\{ \frac{1}{{K_p}'} \exp\left[ -\frac{1}{\beta} \sum_{j \in \mathcal{N}_k} w_i |\zeta - Z_j| \right] \right\} \\
&= \arg\max_{\zeta} \left\{ \ln\left( \frac{1}{{K_p}'} \right) - \frac{1}{\beta} \sum_{j \in \mathcal{N}_k} w_i |\zeta - Z_j| \right\} \\
&= \arg\max_{\zeta} \left\{ -\frac{1}{\beta} \sum_{j \in \mathcal{N}_k} w_i |\zeta - Z_j| \right\} \\
&= \arg\max_{\zeta} \left\{ - \sum_{j \in \mathcal{N}_k} w_i |\zeta - Z_j| \right\} \\
&= \arg\min_{\zeta} \left\{ \sum_{j \in \mathcal{N}_k} w_i |\zeta - Z_j| \right\} \qquad (23)
\end{aligned}$$

# ESTIMATION - AVMRF & WOS Filt. 3

- Weighted Order Statistic (WOS) Filters
  - an apparent detour – weighted median filters
- Typical formulation
  - data value $\{x_1, \ldots, x_i, \ldots, x_n\}$
  - weights $\{w_1, \ldots, w_i, \ldots, w_n\}$, $w_i \in \{1, 2, \ldots\}$
  - $x_{wos} = \text{med}\{x_1 \diamond w_1, \ldots, x_i \diamond w_i, \ldots, x_n \diamond w_n\}$
  - $a \diamond b$ – replicate $a$, $b$ times
- Problem: what if . . .
  - list has an even number of entries?
  - weights are non-integer values?
- More general formulation

$$x_{wos} = \arg\min_x \sum_{i=1}^{K} w_i |x - x_i| \qquad (24)$$

# ESTIMATION - AVMRF & WOS Filt. 4

- Compare:
  - WOS filtering equation

$$x_{wos} = \arg\min_{x} \sum_{i=1}^{K} w_i |x - x_i| \tag{25}$$

  - MAP prediction of a pixel given its neighbors

$$\hat{Z}_k = \arg\min_{\zeta} \sum_{j \in \mathcal{N}_k} w_i |\zeta - Z_j| \tag{26}$$

- Identical form!

- AVMRF MAP prediction $\equiv$ WOS prediction

# ESTIMATION - AVMRF MAP
# Prediction

- Graphical example 1

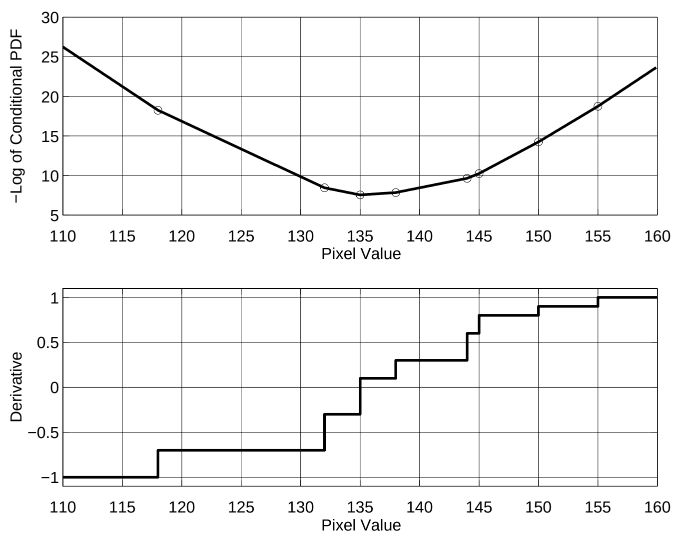



Figure 13: Negative logarithm of the conditional PDF of a pixel given its neighbors (above) and its derivative (below), in an instance where the optimal prediction is a unique value.

# ESTIMATION - AVMRF MAP
# Prediction

- Graphical example 2

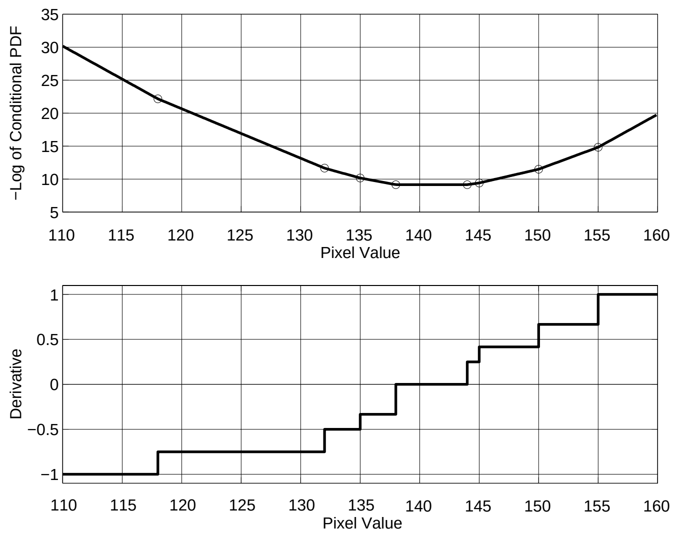



Figure 14: Negative logarithm of the conditional PDF of a pixel given its neighbors (above) and its derivative (below), in an instance where the optimal prediction is an interval of values.

# ESTIMATION - MPEE 1

- Question:
  - *given a single image observation, is it possible to estimate the clique weight parameters of the absolute value MRF that best fits the data ?*
- Scenario:
  - model clique weight parameters unknown
  - realization data available
  - $\rightarrow$ prediction of pixel given neighbors possible
  - "correct" answer known since pixel values known
- Proposal:
  - *seek the parameters of the weighted median predictor which tends to minimize the error between the weighted median prediction of a pixel and the actual value of that pixel*
  - minimize error between MAP prediction of a pixel given its neighbors and the known value of that pixel from the realization
  - $\rightarrow$ Minimization of Prediction Error (MPE)

# ESTIMATION - MPEE 2

- Minimization of Prediction Error Estimation (MPEE)
  - use of MPE principle for estimation
- Inspiration:
  - optimal weight selection for WOS filters
  - substantial body of literature
  - cross validation
- Application:
  - AVMRF clique weight estimation
  - with hopes of extending to other models

# ESTIMATION - MPEE 3

- MPEE for clique weights (formal)
  - $Z$ - known realization image vector
  - $\hat{Z}$ - predicted image vector (fn of model params)
  - seek clique weights $W$, so write just $\hat{Z}(W)$
  - MPEE optimal weights $W_{opt}$ found by:

$$W_{opt} = \arg\min_{W} \left\{ \mathcal{E} \left[ \mathcal{C} \left( \hat{Z}(W) - Z \right) \right] \right\} \quad (27)$$

  - $\mathcal{E}$ - the expectation operation
  - $\mathcal{C}$ - cost fn applied to the pixel prediction error

# ESTIMATION - MPEE Cost Fn's 1

- Choice of the prediction error cost function
  - how do we choose $\mathcal{C}$ ?
  - much freedom
  - considered cost function on the basis of
    1. their equivalence with existing methods
    2. attractiveness for our particular problem

# ESTIMATION - MPEE Cost Fn's 2

- Absolute value cost
  - cost functional

$$\mathcal{C}(\epsilon) = |\epsilon| \tag{28}$$

  - optimization:

$$W_{opt} = \arg\min_{W} \left\{ \mathcal{E}\left( \left| \hat{Z}(W) - Z \right| \right) \right\} \tag{29}$$

  - assuming ergodicity, replace ensemble average with the spatial average over the point set $\mathcal{S}$

$$W_{opt} = \arg\min_{W} \left\{ \frac{1}{|\mathcal{S}|} \sum_{i \in \mathcal{S}} \left| \hat{Z}_i(W) - Z_i \right| \right\} \tag{30}$$

$\hat{Z}_i(W)$ - MAP prediction of $i^{\text{th}}$ pixel value
$|\mathcal{S}|$ - cardinality of point set $\mathcal{S}$

  - minimizes the mean absolute error (MMAE)

# ESTIMATION - MPEE Cost Fn's 3

- Square cost
  - cost functional

$$\mathcal{C}(\epsilon) = |\epsilon|^2. \tag{31}$$

  - optimization:

$$W_{opt} = \arg\min_{W} \left\{ \mathcal{E}\left( \left| \hat{Z}(W) - Z \right|^2 \right) \right\} \tag{32}$$

  - under ergodicity assumptions

$$W_{opt} = \arg\min_{W} \left\{ \frac{1}{|\mathcal{S}|} \sum_{i \in \mathcal{S}} \left| \hat{Z}_i(W) - Z_i \right|^2 \right\}. \tag{33}$$

  - minimizes the mean of the squared error (MMSE)

# ESTIMATION - MPEE Cost Fn's 4

- Zero/One cost function
  - cost functional

$$\mathcal{C}(\epsilon) = \begin{cases} 0, & |\epsilon| \leq \alpha, \\ 1, & |\epsilon| > \alpha. \end{cases} \tag{34}$$

  - optimization:

$$W_{opt} = \arg\min_{W} \left\{ \mathcal{E}\left( \left\{ \begin{array}{ll} 0, & |\hat{Z}(W) - Z| \leq \alpha, \\ 1, & |\hat{Z}(W) - Z| > \alpha. \end{array} \right\} \right) \right\}, \tag{35}$$

  - under ergodicity assumptions

$$W_{opt} = \arg\min_{W} \left\{ \frac{1}{|\mathcal{S}|} \sum_{i \in \mathcal{S}} \left\{ \begin{array}{ll} 0, & |\hat{Z}_i(W) - Z_i| \leq \alpha, \\ 1, & |\hat{Z}_i(W) - Z_i| > \alpha. \end{array} \right\} \right\} \tag{36}$$

# ESTIMATION - MPEE Cost Fn's 5

- Zero/One cost function ctnd . . .
  - "MAP" pixel predictor is only the actual MAP predictor if the true clique weights are used for prediction
  - We are in effect formulating predictors with the equivalent form as the MAP pixel predictor and testing their performance using various cost functions
  - MAP pixel predictor is exactly that predictor which yields minimum risk using the zero/one cost function for $\alpha \to 0$
  - Therefore if we find a consistent estimator of minimum risk using the zero/one cost function as $\alpha \to 0$, we have found the MAP estimator and therefore the exact clique weights
  - We therefore seek a consistent estimator for $\alpha \to 0$
  - Unfortunately it is not possible to achieve this since no matter how much we increase the data size, as we make $\alpha \to 0$, the variance of the prediction error does not tend to zero.

# ESTIMATION - MPEE Cost Fn's 6

- Problem with zero/one cost
  - derivative is zero almost everywhere
  - $\rightarrow$ problems in MPE optmization
  - use similar, but differentiable functions

- Piecewise linear cost

$$\mathcal{C}(\epsilon) = \begin{cases} \frac{1}{\alpha}|\epsilon|, & |\epsilon| \leq \alpha, \\ \beta|\epsilon| + (1 - \alpha\beta), & |\epsilon| > \alpha. \end{cases} \quad (37)$$

- Power cost

$$W_{opt} = \arg\min_{W} \left\{ \mathcal{E}\left( \left| \hat{Z}(W) - Z \right|^{\gamma} \right) \right\} \quad (38)$$

# ESTIMATION - MPEE Cost Fn's 7

- Graphical examples of cost functions

  1. absolute value cost
  2. squared cost
  3. zero/one cost ($\alpha = 0.1$)
  4. power function cost ($\gamma = 0.1$)
  5. piecewise linear cost ($\alpha = 0.1, \beta = 0.1$)

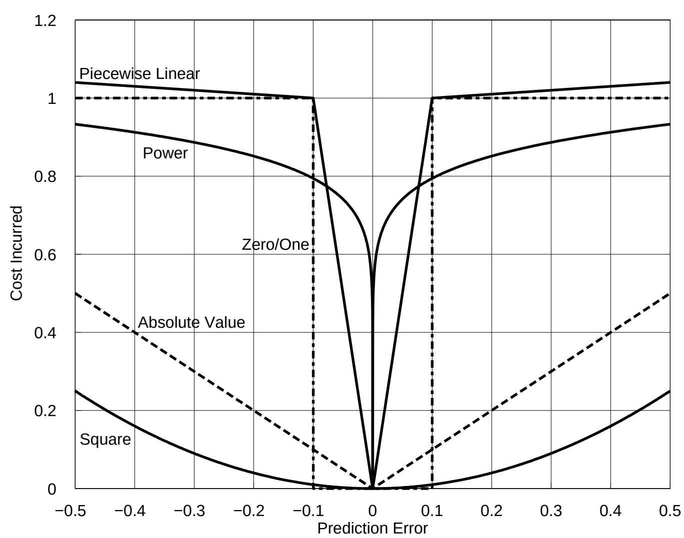



Figure 15: Cost functionals used for penalization of prediction error in MPE estimation.

# ESTIMATION - MPEE for AVMRF 1

- Application of MPEE
  - MPEE equation

$$W_{opt} = \arg\min_{W} \left\{ \mathcal{E}\left[ \mathcal{C}\left( \hat{Z}(W) - Z \right) \right] \right\}. \quad (39)$$

  - Assume ergodicity
    Expectation $\equiv$ Spatial average

$$W_{opt} = \arg\min_{W} \left\{ \frac{1}{|\mathcal{S}|} \sum_{i \in \mathcal{S}} \mathcal{C}\left( \hat{Z}_i(W) - Z_i \right) \right\}. \quad (40)$$

  - Define *risk* $\mathcal{R}$ similar to expected cost

$$\mathcal{R} \triangleq \frac{1}{|\mathcal{S}|} \sum_{i \in \mathcal{S}} \mathcal{C}\left( \hat{Z}_i(W) - Z_i \right). \quad (41)$$

  - MPEE: find $W_{opt}$ for which $\mathcal{R}$ is minimum

# ESTIMATION - MPEE for AVMRF 2

- How to minimize $\mathcal{R}$ given an AVMRF realization?
  - exhaustively search entire $W$ space?
  - not feasible

- Examine behavior of $\mathcal{R}(W)$
  - $\mathcal{R}$ is weighted summation of penalized prediction errors for each individual image pixel
  - understand effect of of $W$ on individual prediction
    $\rightarrow$ understand effect on $\mathcal{R}$

- MAP pixel prediction behavior as a function of $W$
  - tutorial example - $1^{\mathrm{st}}$ order predictor
  - prediction based on 4-neighbors

# ESTIMATION - MPEE for AVMRF 3

- Tutorial example 1:
  - prediction is interval of values

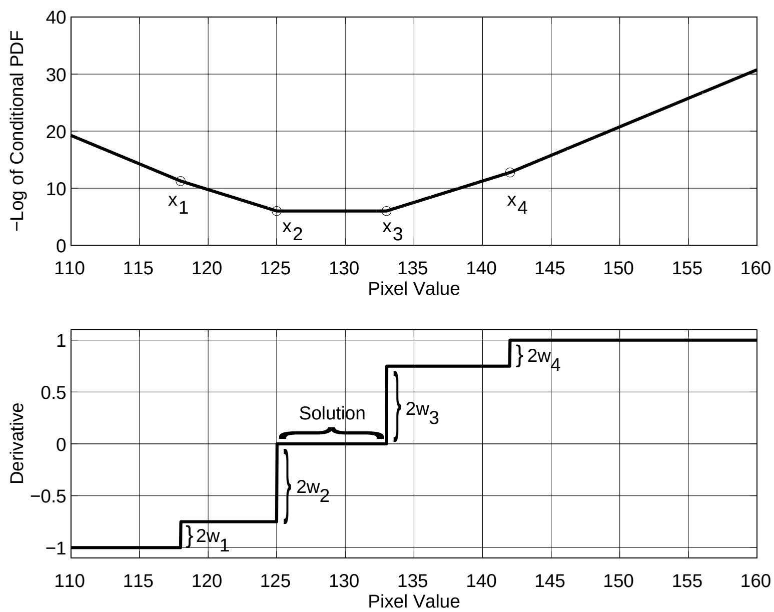



Figure 16: -log of the cond PDF of pixel given neighbors and derivative.

- how is prediction affected by change of $W$?
- prediction unchanged for $0 \leq w_1 < 0.5$
- for $w_1 = 0.5$ prediction interval is $x_1$ to $x_4$
- varying $w_i$ will not affect prediction unless $w_i$ changes to, or from, $0.5$.
- this is not the only interesting configuration

# ESTIMATION - MPEE for AVMRF 4

- Tutorial example 2:
  - prediction is unique

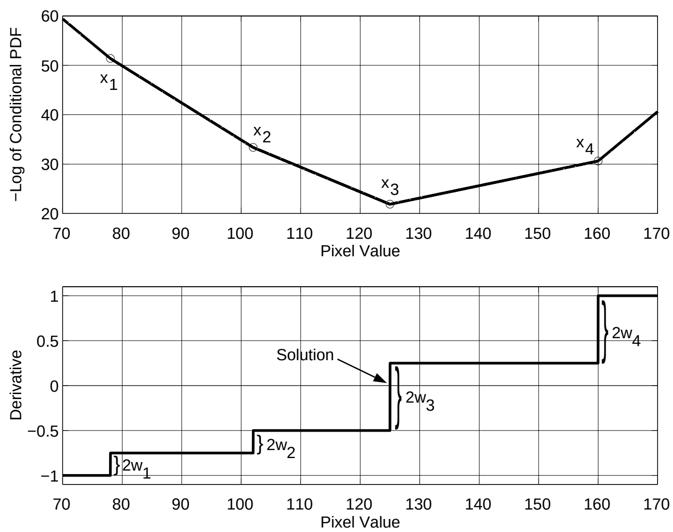



Figure 17: -log of the cond PDF of pixel given neighbors and derivative.

- how is prediction affected by change of $W$?
- prediction is interval from $x_3$ to $x_4$ if $w_1 = 0$
- prediction is $x_3$ for $0 < w_1 < 0.25$
- prediction is interval from $x_2$ and $x_3$ if $w_1 = 0.25$
- prediction is $x_2$ if $0.25 < w_1 < 0.5$
- prediction is interval from $x_1$ to $x_2$ if $w_1 = 0.5$

# ESTIMATION - MPEE for AVMRF 5

- What can we conclude for $1^{\text{st}}$ order predictor?

| Weight Value | Solution Behavior |
|---|---|
| $w = 0$ | Possible change |
| $0 < w < 0.25$ | Constant |
| $w = 0.25$ | Possible change |
| $0.25 < w < 0.5$ | Constant |
| $w = 0.5$ | Possible change |

Table 1: Behavior of pixel prediction as fn of indep clique weight $w$.

- prediction does not change for all values of $w$
- $\rightarrow$ need not examine prediction for all $w$!
- need only examine prediction at:
  1. $w = 0, 0.25, 0.5$
  2. one point in $0 < w < 0.25$
  3. one point in $0.25 < w < 0.5$
- conveniently choose $w = \frac{k}{2N}$
  $N$ is number of neighbors
  $k$ is integer $0 \leq k \leq N$

# ESTIMATION - MPEE for AVMRF 6

- Risk behavior as a function of clique weight
  - risk is weighted avg of penalized prediction errors
  - exhibits same constant valued regions as observed for individual predictions
  - example:

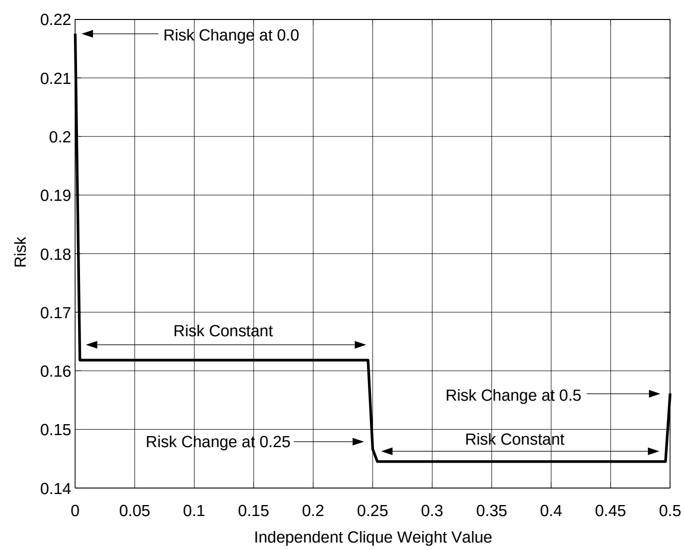



Figure 18: Risk as a fn of indep clique weight for $1^{\mathrm{st}}$ order predictor.

# ESTIMATION - MPEE for AVMRF 7

- Implications of behavior of risk
  - Quantization effect
    - in constant valued regions of $\mathcal{R}(W)$ all choices of weights which result in that risk are equally good choices
    - all such choices are optimal in MPE sense
    - there are a "range" of optimal solutions
    - we call this the "*quantization effect*"
    - limits precision of clique weight estimate
  - Grid searching of $W$-space for minimization of $\mathcal{R}$
    - $\mathcal{R}(W)$ exhibits constant valued regions
    - $\rightarrow$ need only evaluate $\mathcal{R}(W)$ at $w = \frac{k}{2N}$
    - searching is on point grid, not entire $W$ space
    - $\rightarrow$ minimization of $\mathcal{R}(W)$ feasible and efficient
  - Note:
    - if minimum risk $w = \frac{k}{2N}$, has $k$ odd (weight is center of interval of optimal weight values), then this solution is *not unique* . . .
    - a range of $w$ values satisfy the MPE criterion
    - $0 \leq \frac{k-1}{2N} < w < \frac{k+1}{2N} \leq \frac{1}{2}$.

* simply a restatement of the quantization effect

# ESTIMATION - MPEE for AVMRF 8

- Generalization to higher order predictors
  - $1^{\text{st}}$ order predictor . . .
    grid searching using indep weight $w = \frac{k}{2N}$
  - what about higher order predictors?
  - number of indep weights increases with order
  - claim:
    * $w = \frac{k}{2N}$ searching valid in higher dimensions
    * grid searching extends to higher dimensions
  - proof:
    * similar arguments to those used for $1^{\text{st}}$ order
    * formal proof: by induction?

# ESTIMATION - MPEE for AVMRF 9

- Problem with AVMRF MPEE of $W$
  - all appears fine . . .
  - implement clique weight MPEE for AVMRF
  - use simulation test images (known MRF params)
  - examine results (thesis: appendix b)
    * input simulation clique weights vary significantly
    * estimated weights are always $0$ or $0.5$
    * estimates are *qualitatively* correct . . .
    * but not *quantitatively*
    * estimates not within quantization effect bounds
  - what's the problem?

- Clues to the cause of the problem
  - all weight assigned to a single symmetry pair, irrespective of input MRF weights
  - risk MPEE depends on individual predictions
  - $\rightarrow$ examine individual predictions carefully?

# ESTIMATION - MPEE for AVMRF 10

- Effect of individual predictions on $\mathcal{R}$
  - MPE: minimize risk $\equiv$ minimizing prediction error
  - behavior of prediction error for $w = 0$ or $w = 0.5$?
  - $\rightarrow$ prediction is almost always an interval
  - if interval contains true value, prediction error is zero
  - if prediction is unique, prediction error is almost always $> 0$
  - prediction error for interval $\leq$ prediction error for unique prediction?
  - mostly true
  - to minimize $\mathcal{R}$, choose weights which lead to many interval predictions
    * likely to contain the true value
    * $\rightarrow$ zero prediction error
    * $\rightarrow$ zero contribution to $\mathcal{R}$
  - which weights lead to interval solutions?
    * surprise . . .
    * $w = 0$ or $w = 0.5$ !

# ESTIMATION - MPEE for AVMRF 11

- AVMRF MPEE problem explained
  - interval type predictions are lower risk
  - $\rightarrow$ these weights are MPE optimal
  - $\rightarrow$ degenerate estimate behavior

- A solution – ad-hoc
  - degenerate behavior results from interval solutions
  - remove possibility of interval type solutions!
  - use mid-point of interval instead of interval itself
  - criticism:
    - completely ad-hoc and somewhat arbitrary
    - no obvious claim to optimality
  - but does this really work?
  - yes!
  - thesis: tables 4.2 – 4.5

# ESTIMATION - MPEE for AVMRF 12

- AVMRF MPEE for higher order predictors
  - $\uparrow$ order of predictor $\rightarrow \uparrow$ precision of estimate
  - $\uparrow$ order $\rightarrow \uparrow N \rightarrow$ smaller grid search spacing $\frac{1}{2N}$.
  - $\uparrow$ order allows modeling of more complex MRF's
  - $\uparrow$ order does not give information where none previously existed
    * e.g. MPEE using second order predictor on first order MRF gives no more information than first order predictor MPEE

# ESTIMATION - GGMRF & GWOS Filters

- Relationship between GGMRF and GWOS Filters
- Form of GGMRF

$$\Pr(\mathbf{Z}) = \frac{1}{K_p} \exp\left[-\frac{1}{\beta} \sum_{c \in \mathcal{C}_2} w_i \left|\mathcal{A}_c(\mathbf{Z})\right|^p\right] \qquad (42)$$

  - $\rho_\alpha(\cdot) = |\cdot|$, $i$ numbers unique types of pair cliques
  - $\mathcal{C}_2$ is the set of pair cliques

- MAP estimate of pixel given neighbors
  - *maximize* $\Pr(Z_k|\mathbf{Z} \setminus Z_k)$
  - apply Markov property of PDF
  - $\Pr(Z_k|\mathbf{Z} \setminus Z_k) = \Pr(Z_k|Z_j)$ for $j \in \mathcal{N}_k$

$$\Pr(Z_k|\mathbf{Z}\setminus Z_k) = \frac{1}{K_p{}'} \exp\left[-\frac{1}{\beta} \sum_{j \in \mathcal{N}_k} w_i |Z_k - Z_j|^p\right] \qquad (43)$$

- $K_p{}'$ is a new normalization constant

# ESTIMATION - GGMRF & GWOS Filters 2

- MAP estimate of pixel given neighbors ctnd . . .
  - find MAP estimate $\hat{Z}_k$, of $k^{\text{th}}$ pixel's value

$$
\begin{aligned}
\hat{Z}_k &= \arg\max_{\zeta} \left\{ \frac{1}{{K_p}'} \exp\left[ -\frac{1}{\beta} \sum_{j \in \mathcal{N}_k} w_i |\zeta - Z_j|^p \right] \right\} \\
&= \arg\max_{\zeta} \left\{ \ln\left( \frac{1}{{K_p}'} \right) - \frac{1}{\beta} \sum_{j \in \mathcal{N}_k} w_i |\zeta - Z_j|^p \right\} \\
&= \arg\max_{\zeta} \left\{ -\frac{1}{\beta} \sum_{j \in \mathcal{N}_k} w_i |\zeta - Z_j|^p \right\} \\
&= \arg\max_{\zeta} \left\{ - \sum_{j \in \mathcal{N}_k} w_i |\zeta - Z_j|^p \right\} \\
&= \arg\min_{\zeta} \left\{ \sum_{j \in \mathcal{N}_k} w_i |\zeta - Z_j|^p \right\} \qquad (44)
\end{aligned}
$$

# ESTIMATION - GGMRF & GWOS Filters 3

- Generalized Weighted Order Statistic (GWOS) Filters
  - our convenient extension of the WOS filter

- Formulation
  - GWOS filter

$$x_{gwos} = \arg\min_x \sum_{i=1}^{K} w_i |x - x_i|^p \tag{45}$$

  - WOS filter

$$x_{wos} = \arg\min_x \sum_{i=1}^{K} w_i |x - x_i| \tag{46}$$

  - GWOS filter $\equiv$ WOS filter if $p = 1$

# ESTIMATION - GGMRF & GWOS
# Filters 4

- MAP prediction of a pixel given its neighbors for GGMRF takes the form of a GWOS prediction
- Compare:
  - GWOS filtering equation

$$x_{gwos} = \arg\min_{x} \sum_{i=1}^{K} w_i |x - x_i|^p \tag{47}$$

  - MAP prediction of a pixel given its neighbors

$$\hat{Z}_k = \arg\min_{\zeta} \sum_{j \in \mathcal{N}_k} w_i |\zeta - Z_j|^p \tag{48}$$

- Identical form by design!
- GGMRF MAP prediction $\equiv$ GWOS prediction

# ESTIMATION - GGMRF MAP
# Prediction 1

- Graphical example 1

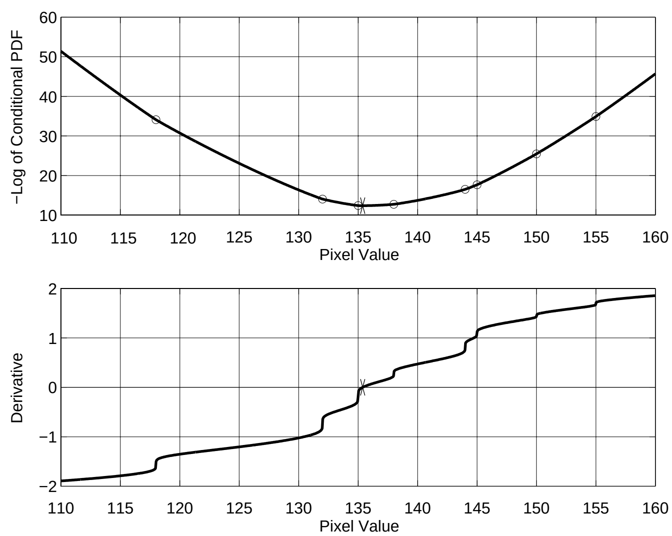



Figure 19: Negative logarithm of the conditional PDF of a pixel given its neighbors (above) and its derivative (below) for a GGMRF.

# ESTIMATION - GGMRF MAP
# Prediction 2

- Graphical example 2

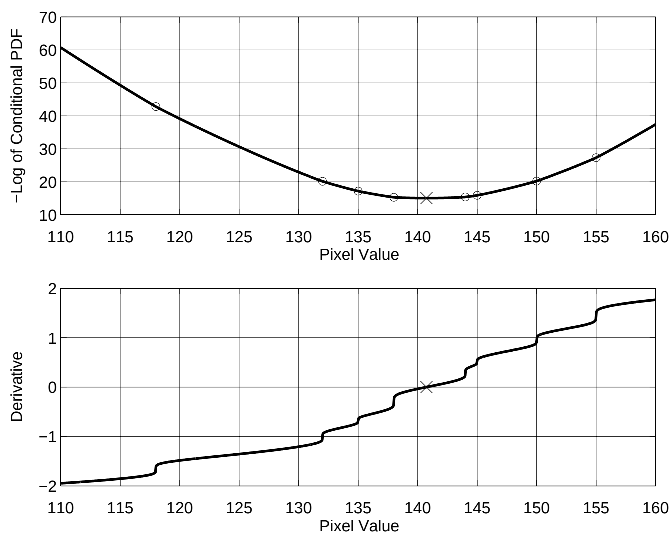



Figure 20: Negative logarithm of the conditional PDF of a pixel given its neighbors (above) and its derivative (below) for a GGMRF. The same neighborhood system led to a non-unique prediction when the data was assumed to be from an AVMRF.

# ESTIMATION - GGMRF MAP
# Prediction 3

- AVMRF MPEE Problem revisited
  - degenerate MPEE results for AVMRF
  - cause: non-uniqueness predictions
  - proposal:
    * when performing MPEE estimation for AVMRF,
      use GGMRF MAP predictor for $p \to 1$
    * interval solutions do not exist
    * $\to$ expect to avoid degenerate behavior,
      while remaining faithful to $p = 1$ model
    * theoretically, more sound than ealier solution
  - does this really work?
  - yes!
  - thesis: tables 4.6 – 4.9
  - interesting observation:
    * results for $p \to 1$ GGMRF MPEE almost identical to those using the ad-hoc method
    * sometimes ad-hoc solutions really *do* work!

# ESTIMATION - MPEE for GGMRF 1

- Application of MPEE
  - Estimate $W$ by MPE principle

$$W_{opt} = \arg\min_{W} \left\{ \mathcal{E}\left[ \mathcal{C}\left( \hat{Z}(W) - Z \right) \right] \right\} \quad (49)$$

  - $\hat{Z}(W)$ is a GWOS predictor
  - MPE-optimal $W$ has minimum risk

$$\mathcal{R} \triangleq \frac{1}{|\mathcal{S}|} \sum_{i \in \mathcal{S}} \mathcal{C}\left( \hat{Z}_i(W) - Z_i \right). \quad (50)$$

  - How do we find $W$ for which $\mathcal{R}$ is minimum?
    * previously relied on convenient characteristics of $\mathcal{R}(W)$
    * no such advantage for GGMRF predictor
    * $\mathcal{R}$ has no convenient dependence on $W$
    * what do we observe for $\mathcal{R}(W)$?

# ESTIMATION - MPEE for GGMRF 2

- Example: GGMRF MPEE $\mathcal{R}(W)$

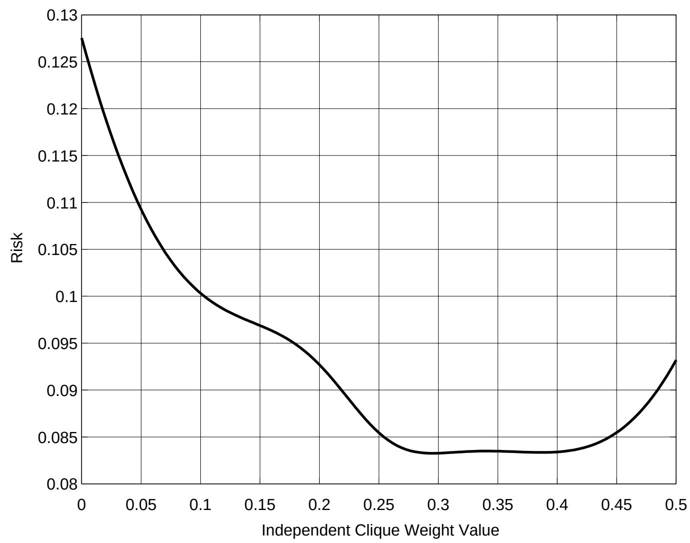



Figure 21: Risk as a function of independent clique weight for a $1^{\mathrm{st}}$ order GGMRF predictor model with $p = 1.3$. The input realization is a GGMRF with clique weight vector $\frac{1}{6}[0\ 2\ 0\ 1]^T$ and $p = 1.3$. The square of the prediction error penalty was used.

# ESTIMATION - MPEE for GGMRF 3

- Gradient Descent minimization of $\mathcal{R}$
  - example suggests GD optimization is feasible
  - how do we do it?
    * require expression for gradient!
    * thesis: appendix A - long and unexciting
    * GD algorithm:

      1. **Initialize**
         Set $M = 0$
         Set $W^M$ (initial guess for minimum risk $W$)
         Find the expression for $[\nabla_{\mathcal{R}}]_k = \frac{\partial \mathcal{R}}{\partial W_k}$
      2. **Update**
         $W^{M+1} = W^M - \alpha \cdot \nabla_{\mathcal{R}}(W^M)$
         $M \leftarrow M + 1$
      3. **Convergence Check**
         if $\|\nabla_{\mathcal{R}}(W^M)\| < \tau$, stop, else back to 2.

    * $\alpha$ – step size $> 0$, chosen experimentally
    * $\tau$ – convergence tolerance, chosen experimentally
    * $\|\cdot\|$ – any norm

# ESTIMATION - MPEE for GGMRF 4

- Gradient Descent minimization example

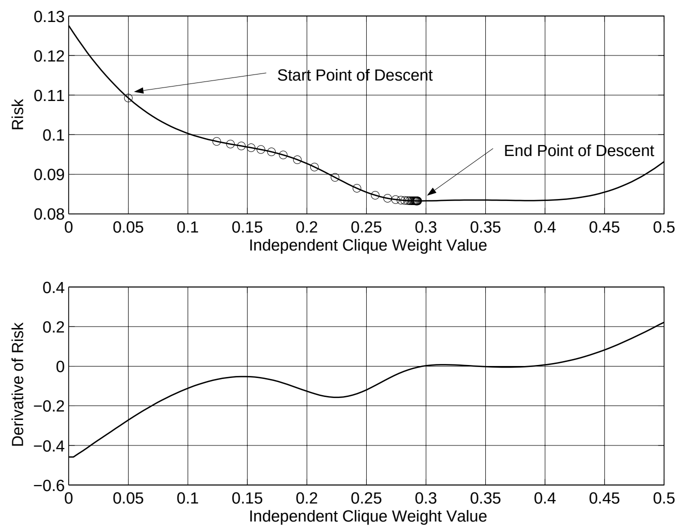



Figure 22: Risk function and the trajectory of a gradient descent search for a minimum of the function. Shown in lower frame is the derivative of the risk.

# ESTIMATION - MPEE for GGMRF 5

- GD *appears* to be viable for minimization of $\mathcal{R}(W)$
  - Figure 21 has 2 local minima!
  - Are all $\mathcal{R}(W)$ functions even smooth?

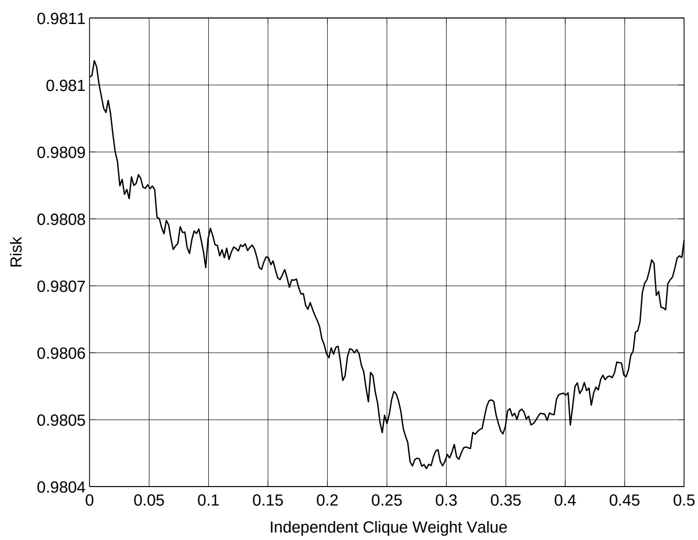



Figure 23: Risk function for which gradient descent optimization is ineffective.

# ESTIMATION - MPEE for GGMRF 6

- What can we do with GD now?
  - choose a good initial guess
  - apply GD to smooth $\mathcal{R}(W)$

- Performance of GD MPEE
  - initial guess is result of grid searching MPEE
  - apply GD to $|\epsilon|^2$ penalized predictions (smooth)
  - results – thesis: table 4.10
  - works, under favorable conditions
  - is GD minimum the global minimum?

- GD unsatisfactory
  - unreliable

- Alternatives to GD?

# ESTIMATION - MPEE for GGMRF 7

- Problem with GD
  - GD gets stuck in local minima
  - Is there a method to avoid this?
  - Yes - don't always descend $\mathcal{R}$
  - Optimization by "simulated annealing" is such a method

- Annealing in a physical system:
  - What is annealing?
    * thermal process whereby a highly ordered, low energy state is a achieved in a solid lattice, after melting and very careful cooling back to the solid state
    * brings system to lowest energy state

- Metropolis algorithm
  - Computational procedure for simulating the evolution of a lattice system in a heat bath to *thermal equilibrium*
  - At equilibrium, probability of states of the lattice follow a Boltzmann, or Gibbs, distribution

# ESTIMATION - MPEE for GGMRF 8

- Metropolis algorithm
  1. Choose an initial state $S_0$.
  2. Given $S_i$, propose $S_{i+1}$.
  3. Accept $S_{i+1}$ with prob, $\min\left\{1, \exp\left(-\frac{(E_{i+1}-E_i)}{k_B T}\right)\right\}$
  4. Return to step 2

  – $E_i$, $E_{i+1}$ are energies of states $S_i$, $S_{i+1}$
  – $k_B$ is the Boltzmann constant
  – $T$ is the temperature of the heat bath

  – running Metropolis algorithm for sufficient duration ensures lattice system reaches *thermal equilibrium* with heat bath at temperature $T$.
  – at equilibrium, states $\mathbf{X}$, of lattice system are Boltzmann/Gibbs distributed
  – PDF of lattice states at thermal equilibrium

$$\Pr(\mathbf{X} = S_i) = \frac{1}{K_p(T)} \exp\left(\frac{-E_i}{k_B T}\right). \qquad (51)$$

$E_i$ - energy of state $S_i$
$K_p(T)$ - partition function

# ESTIMATION - MPEE for GGMRF 9

- Annealing revealed . . .
  - *Carefully* lower temperature $T$
  - Lattice system must reach thermal equilibrium $\forall T$
  - Gibbs distribution becomes increasingly "peaked"
  - Only lower energy states have significant prob
  - Eventually, lattice "freezes" in lowest energy state
  - *Energy of lattice system is minimized*

- Optimization by simulated annealing
  - Kirkpatrick, Gellat and Vecchi
  - Apply annealing to *optimization* problems
  - Identify the following equivalences:
    1. state space of a physical system
       $\leftrightarrow$ parameter space of an optimization problem
    2. energy of the physical system
       $\leftrightarrow$ risk function of the optimization problem
  - Annealing isolates:
    1. *minimum energy state* in a physical system

2. *minimum risk state* in the optimization problem

# ESTIMATION - MPEE for GGMRF 10

- Distinguishing features of simulated annealing
  - risk is *not* monotonically decreased
  - always prob that *higher* risk state is accepted
  - simulated annealing = "stochastic descent"
  - won't "get stuck" in local minima

- Complications
  - rate at which to reduce the $T$?
    * if $T \downarrow$ too quickly,
      system $\rightarrow$ only *locally* optimal state
    * if $T \downarrow$ too slowly,
      anneal is computationally inefficient
    * *annealing schedule* must be selected
    * schedules exist that guarantee global minimum
    * problem: too slow for practical use
    * schedules selected according to problem

# ESTIMATION - MPEE for GGMRF 11

- Graphical example of annealing MPEE

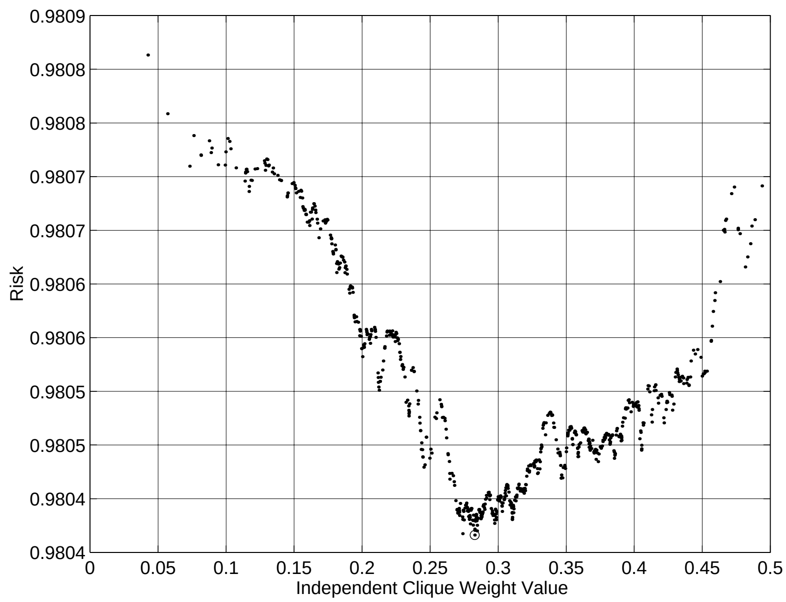



Figure 24: Weight values accepted in a simulated annealing optimization of a complex risk functional.

# ESTIMATION - MPEE for GGMRF 12

- Performance of annealing MPEE
  - thesis: table 4.11
  - very effective, even for highly non-convex $\mathcal{R}$
  - problems:
    - *very* computationally costly
    - selection of annealing schedule is an art
    - graphical example of failed annealing MPEE
      - thesis: figures 4.15 and 4.16
      - temperature parameter dropped too quickly
      - system "freezes" in risk local minimum

- MPEE for other classes of MRF
  - we applied MPEE to GGMRF
  - applicable to other models
  - example: Huber MRF
    - predictor would different
    - MPE principle the same

# CONCLUSIONS - Research Contribution

- Proposed and implemented a scheme for estimation of the clique weight parameters of the generalized Gaussian Markov random field, using the technique we call *minimization of prediction error estimation* or MPEE.

- Optimal clique weight parameters for the GGMRF are estimated by minimizing the weighted error or *risk* between a prediction, based on observed data, and the known realization data itself.

- The predictor model derives from the MAP estimate of an image pixel given its neighbors for the given MRF model, and is shown to be equivalent in form to what we call a *generalized weighted order statistic* (GWOS) predictor

- Performance tests of the MPEE algorithm were performed, with favorable results.

# CONCLUSIONS - Future Investigation 1

- **MRF Simulation**
  - Theoretical Issues
    - Optimally efficient MRF simulation
      - iterative simulation is costly
      - sample rejection by Metropolis algorithm
      - scope for work in “optimal” methods
    - Convergence testing for MRF simulation
      - theoretical framework for convergence testing
      - what class of parameters estimated from chain guarantee convergence of the chain itself?
  - Computational Issues
    - Performance enhancement
      - large images
      - large neighborhoods
      - simulation is part of estimation algorithm

# CONCLUSIONS - Future Investigation 2

- **MPE Estimation**
  - Theoretical Issues
    - Formal proof of generalization of grid searching to higher dimensions
    - Choice of optimal cost functions in risk computation
      - zero/one cost MPEE $\rightarrow$ *exact* MRF $W$
      - problem: only asymptotically true
      - for typically sized data set, zero/one cost with $\alpha \rightarrow 0$ results in most predictions being penalized, and risk functionals with high variance which are difficult to optimize.
      - optimal cost fn for typical sized data sets?
    - Determining optimality criteria for cost functions
      - desirable statistical properties of estimator
      - conveniently optimized risk functionals
    - Statistical behavior of MPE estimators
      - NB: theoretical analysis required!
    - Extending MPEE to other MRF's and other parameters

# CONCLUSIONS - Future Investigation 3

- **MPE Estimation**
  - Computational Issues
    - Alternative algorithms for risk minimization
      - cheaper optimization methods?
    - Analysis of computational cost of MPEE
      - useful for comparison and improvement

# CONCLUSIONS - Future Investigation 4

- **Applications**
  - Motivation for this work:
    * belief that the use of more realistic image models as *a-priori* knowledge in image reconstruction and restoration problems will lead to improved performance of these methods
    * now in a position to test this belief
    * test to determine whether performance gains are evident when utilizing MRF's, whose clique weight parameters are determined using the MPEE method from sample solution imagery, as *a-priori* knowledge for Bayesian tomographic reconstruction
    * Previously, the clique weights of the MRF prior were typically chosen on an ad-hoc basis. Naturally, it is our hope that our approach will prove successful.