Lecture 3
Day 1

Note Title 7/3/2011

Weak Smoothness.

Image data suggests that images are piecewise smooth. This lecture describes models that exploit this knowledge for image segmentation and image de-noising / reconstruction.

Simplest, and most commonly used, model is the Total Variation (TV) norm. Rudin, Osher, Fatemi.

Input image $I(x)$, Output image $\omega(x)$. ← L1 norm.

$$E[\omega:I] = \lambda \int_{\mathcal{D}} \{I(x)-\omega(x)\}^2 dx + \int_{\mathcal{D}} |\nabla \omega(x)|\, dx$$

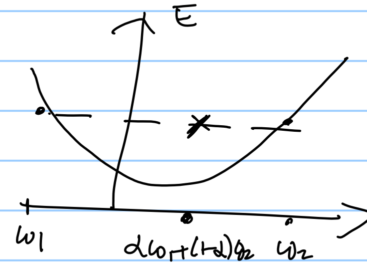
$\mathcal{D}$ 2D-image domain.

Estimate: $\hat{\omega} = \arg\min_{\omega} E[\omega:I]$.

This energy functional $E[\omega:I]$ is convex.

– i.e. $\alpha E[\omega_1:I] + (1-\alpha) E[\omega_2:I] \geq E[\alpha\omega_1 + (1-\alpha)\omega_2 : I]$

for all ω_1, ω_2, $0 \leq \alpha \leq 1$.



Also $E[\omega:I] \geq 0$.

Convexity, and $E[\omega:I] \geq 0$, ensures that $E[\omega:I]$ has only one minimum – i.e. a global minimum.

Therefore we can find the minimum by steepest descent.

$$\frac{d\omega}{dt} = -\frac{\partial E}{\partial \omega} \quad \leftarrow \text{gradient of the energy.}$$

$$\frac{dE}{dt} = \frac{\partial E}{\partial \omega}\frac{d\omega}{dt} = -\left(\frac{\partial E}{\partial \omega}\right)^2 \leq 0$$

$= 0$, at the minimum where $\frac{\partial E}{\partial \omega} = 0$.

In practice, discretize $\omega(t+\Delta) = -\omega(t) - \Delta \frac{\partial E}{\partial \omega}(t)$

Many ways to perform steepest descent. (see. Osher et al.)

Alternative, variational bounding / CCCP.

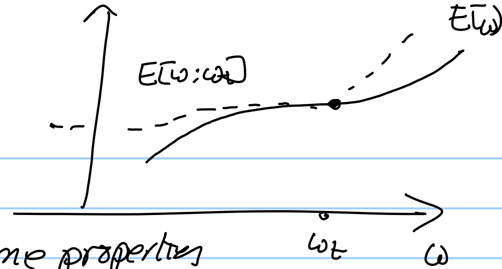
Define. $E[\omega:\omega_t]$ such that $E[\omega_t:\omega_t] = E[\omega_t]$

$E[\omega:\omega_t] \geq E[\omega]$, for all ω

Here ω_t is current state of algorithm.

Page 2.

Variational Boundary:



Select $\omega_{t+1} = \text{ARG MIN } E[\omega:\omega_t]$

construct $E[\omega:\omega_{t+1}]$ with same properties

repeat.

This algorithm is guaranteed to decrease the energy $E[\omega]$

Special Case: Decompose

$$E[\omega] = E_{cave}[\omega] + E_{vex}[\omega]$$

concave energy, convex energy

This decomposition can always be done.

By definition of convexity

$$E_{cave}[\omega] \leq E_{cave}[\omega_t] + (\omega - \omega_t) \cdot \frac{\partial E_{cave}}{\partial \omega}$$

Define $E[\omega:\omega_t] = E_{vex}[\omega] + E_{cave}[\omega_t] + (\omega-\omega_t) \cdot \frac{\partial E_{cave}}{\partial \omega}$

Minimize. $E[\omega:\omega_t]$ w.r.t. ω to get ω_{t+1}

$$\frac{\partial}{\partial \omega} E_{vex}[\omega_{t+1}] = -\frac{\partial}{\partial \omega} E_{cave}[\omega_t]$$

This gives a discrete update equation which is guaranteed to decrease the energy at each iteration step and hence converge to the global minimum.

(no need for Δ, unlike steepest descent).

$$P[\omega:I] = \frac{1}{Z} e^{-E[\omega:I]}$$

Gibbs distribution.

small energy $\leftrightarrow$ large probability

large energy $\leftrightarrow$ small probability.

Data Term: $P[\omega:I] = \frac{1}{Z_1} e^{-\int_D (I(x)-\omega(x))^2 dx}$ additive Gaussian noise.

Generative Model – Image $I(x) = \omega(x) + \varepsilon(x)$

Prior Term: $P[\omega] = \frac{1}{Z_2} e^{-\int_D |\underline{\nabla}\omega(x)| d\underline{x}}$ images are weakly smooth.

Why L^1 norm $|\underline{\nabla}\omega(x)|$?

It encourages sparsity.

An alternative $\frac{1}{Z_3} e^{-\int_D |\underline{\nabla}\omega(x)|^2 d\underline{x}}$ is a Gaussian distribution

Non-robust. Smooths out edges.

$$f(\omega;I) = (\omega - I)^2 + \lambda|\omega|$$

What is $\hat{\omega}(I)$?

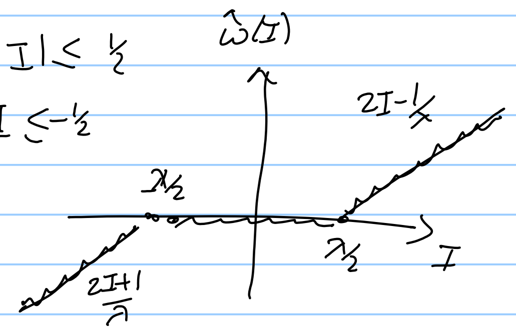
$f_+(\omega;I) = (\omega - I)^2 + \lambda\omega$. For $\omega \geq 0$

$f_-(\omega;I) = (\omega - I)^2 - \lambda\omega$. For $\omega \leq 0$.

$\frac{df_+}{d\omega} = 2(\omega - I) + \lambda$. $\qquad \frac{df_+}{d\omega} = 0 \Rightarrow \hat{\omega}(I) = \frac{2I-1}{\lambda}$, for $\omega \geq 0$

$\frac{df_-}{d\omega} = 2(\omega - I) - \lambda$. $\qquad \frac{df_-}{d\omega} = 0 \Rightarrow \hat{\omega}(I) = \frac{2I+1}{\lambda}$, for $\omega \leq 0$

Hence
$$\hat{\omega}(I) = \frac{2I-1}{\lambda}, \quad I \geq \tfrac{1}{2}$$
$$\hat{\omega}(I) = 0, \quad |I| \leq \tfrac{1}{2}$$
$$\hat{\omega}(I) = \frac{2I+1}{\lambda}, \quad I \leq -\tfrac{1}{2}$$



The use of the L^1 norm $|\omega|$, biases the solution to $\hat{\omega}(I) = 0$ for small $|I|$.

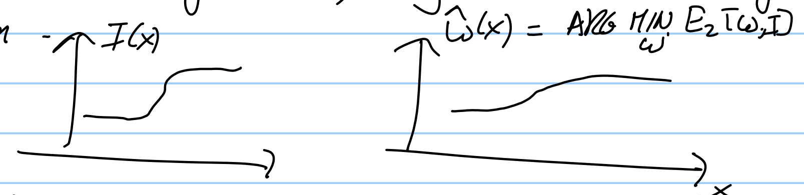
By contrast $\quad f_2(\omega;I) = (\omega - I)^2 + \lambda\omega^2$

has minimum at $\omega - I + \lambda\omega = 0$

$\hat{\omega}(I) = \frac{I}{1+\lambda}$, always smooths I does not force it to 0.

Note: L^1 norm corresponds to an exponential distribution. L^2 norm to a Gaussian distribution

Gaussian distributions are non-robust, very sensitive to outliers. In this image case, they will smooth edges and destroy them -



But the L^1 norm $|\nabla\omega|$ will preserve edges and will bias intensity to flat regions with $\nabla\omega(x) = 0$.

TV-norm state of the art for image denoising until a few years ago.