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 $w(x)$.

$$ E[w; J] = 2 \int_D (I(x) - w(x))^2 \, dx + \int_D |\nabla w(x)| \, dx $$

Estimate: $\hat{w} = \arg\min_w E[w; J]$

This energy functional $E[w; J]$ is **convex**.

- i.e. $2 E[w_1; J] + (1-2)E[w_2; J] > E[w_1 + (1-2)w_2; J]$

  for all $w_1, w_2, \delta \geq 2$.

Also $E[w; J] \geq 0$.

Convexity and $E[w; J] \geq 0$ ensures that $E[w; J]$ has only one minimum.

- i.e. a *global minimum*.

Therefore we can find the minimum by steepest descent.

$$ \frac{dw}{dt} = -\frac{\partial E}{\partial w} = \text{gradient of the energy} $$

$$ \frac{dE}{dt} = \frac{\partial E}{\partial w} \frac{dw}{dt} = -\left(\frac{\partial E}{\partial w}\right)^2 \leq 0 \text{ at the minimum} $$

where $\frac{\partial E}{\partial w} = 0$.

In practice, discretize:

$$ w(t + \Delta t) = w(t) - \Delta t \frac{\partial E}{\partial w} $$

Many ways to perform steepest descent.

(see. Osher et al.)

**Alternative, variational bounding / CCP.**

Define $E[w; w_0]$ such that $E[w_0; w_0] = E[w_0]$.

$$ E[w; w_0] \geq E[w; J] \text{ for all } w $$

Here $w_0$ is current state of algorithm.
Variational Boundary:

Select $w_{t+1} = \arg\min_w E[w; W_t]$

Construct $E[w; W_t]$ with same properties $\log E_w$

Repeat.

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

Special Case: Decompose $E[w]$ into:

$E[w] = E_{\text{data}}[w] + E_{\text{reg}}[w]$ This decomposition can always be done.

By definition of convexity

$E_{\text{data}}[w] \leq E_{\text{data}}[w_{t+1}] + (w - w_t) \cdot \frac{\partial E_{\text{data}}}{\partial w}$

Define $E[w; W_t] = E_{\text{data}}[w] + E_{\text{reg}}[w] + (w - w_t) \cdot \frac{\partial E_{\text{data}}}{\partial w}$

Minimize $E[w; W_t]$ w.r.t. $w$ to get $w_{t+1}$

$\frac{\partial E_{\text{data}}[w_{t+1}]}{\partial w} = -\frac{\partial E_{\text{data}}[w]}{\partial w}$

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 $\Delta$, unlike steepest descent).

$E[w; \mathcal{D}] = 1 \cdot \frac{1}{Z}$ Gibbs distribution

small energy $\implies$ large probability
large energy $\implies$ small probability

Data Term: $P[w; \mathcal{D}] = \frac{1}{Z} e^{-\int_D (I(x) - w(x))^2 dx}$

Gaussian noise.

Generative Model - Image $I(x) = w(x) + \epsilon(x)$

Prior Term: $P[w] = \frac{1}{Z_2} e^{-\int_D |\nabla w(x)| dx}$

$|\nabla w(x)|$ is a Gaussian distribution.

An alternative $\frac{1}{Z_3} e^{-\int_D |\nabla w(x)|^2 dx}$ is a Gaussian distribution.

Non-robust, smooths out edges.
\[ f(w, I) = (w - I)^2 + \lambda w \]

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

\[ f^+(w, I) = (w - I)^2 + \lambda w \quad \text{for } w > 0 \]
\[ f^-(w, I) = (w - I)^2 - \lambda w \quad \text{for } w < 0. \]

\[ \frac{df^+}{dw} = 2(w - I) + \lambda \quad \frac{df^-}{dw} = 2(w - I) - \lambda. \]

Thus,
\[ \hat{\omega}(I) = \begin{cases} \frac{2I - 1}{\lambda} & I \geq \frac{1}{2} \\ \frac{2I + 1}{\lambda} & I \leq \frac{1}{2} \end{cases} \]

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

By contrast, \( f^2(w, I) = (w - I)^2 + \lambda w^2 \) has minimin at \( w = I + \frac{\lambda}{2} \).

\[ \hat{\omega}(I) = \frac{I}{1 + \frac{\lambda}{2}} \quad \text{always smooths } I \]

**Note:** \( L^1 \) norm corresponds to an exponential distribution, \( L^2 \) norm to a Gaussian distribution.

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

But the \( L^1 \) norm \(|w|\) will preserve edges and will bias intensity to flat regions with \( \Delta \hat{\omega}(x) = 0 \).

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