Conference on Stochastic PDEs
Trento, Italy, January 1990

AN SPDE FORMULATION FOR IMAGE SEGMENTATION

Sanjoy K. Mitter
Laboratory for Information and Decision Systems (LIDS)
Massachusetts Institute of Technology
Cambridge, MA 02139

and

Ofer Zeitouni
Department of Electrical Engineering
Technion-Israel Institute of Technology
Haifa 32000, ISRAEL

1 Introduction

In a recent paper [1], it was shown that the optimization problem over $\phi \in W_0^{2,2}(D)$, where $D \subset R^2$ is a nice domain with smooth boundaries,

$$\inf_{\phi}\left(\int_D |\Delta\phi|^2 dx + \int_D \phi \circ \dot{y}(x) - \frac{1}{2}\int_D |\phi|^2 dx\right) \tag{1.1}$$

possesses a solution and can be interpreted as the Maximum a-Posteriori estimator for the solution of the stochastic differential equation

$$\Delta z = n \text{ in } D,\ z\Big|_{\partial D} = 0 \tag{1.2b}$$

observed in the presence of corrupted noise

$$\dot{y} = z + \tilde{n} \tag{1.2a}$$

where $n, \tilde{n}$ are independent white noise processes. For precise definitions of the terms involved, c.f. section 2.

On the other hand, in the context of image segmentation problems, some attention has been generated (cf. [2,5,6]) by the following variational problem:

$$\inf_{\phi,\Gamma} \int_{D\setminus\Gamma} \alpha|\nabla\phi|^2 dx + \int_D (\phi - g)^2 + \beta\ell(\Gamma) \tag{1.3}$$

where $\ell(\Gamma)$ denotes the length of Γ and (1.3) is a free boundary optimization problem.

Various existence and convergence results have been obtained for (1.3) ([2,6]), although no probabilistic interpretation of (1.3) is known to us.

When comparing (1.3) to (1.1), three differences come to mind, i.e.:

(a) $\int_D(\phi - g)^2$ has been replaced by $\int_D \phi \circ \dot{y} - \frac{1}{2}\int_D |\phi|^2 dx$

(b) The gradient term in (1.3) has been replaced by the Laplacian in (1.1).

(c) No segmentation term appears at all in (1.1).

While dealing with (a) is easily understood via the usual Girsanov like transformation, and the reason for (b) has been made explicit in [4], it is not clear how to add segmentation to (1.1) in order to relate it to the framework of (1.3).

In this note, we propose, following [3], a way to do that. Specifically, we construct an appropriate notion of "density" over **closed polygons** and random fields inside the polygons such that, if n denotes the number of straight lines components, one has that maximizing this "density" is equivalent to minimizing

$$\int_{D\setminus\bigcup_{i=1}^n \Gamma_i} \alpha|\Delta\phi|^2 dx + \int_D \phi \circ \dot{y} - \frac{1}{2}\int_D |\phi|^2 dx + 2\sum_{i=1}^n \ell(\Gamma_i) + f(n) \tag{1.4}$$

for an appropriate $f(n)$, over $\phi \in W_0^{2,2}(D\setminus\bigcup_{i=1}^n \Gamma_i)$, n and set of straight lines Γ_i which form (together with ∂D) a polygonal partition of D, $i = 1, \cdots, n$.

The organization of this note is as follows: in Section 2 we recall some results from [3] where a measure on close polygonal partitions of D is constructed. By appropriately defining "densities" for such partitions, and using the density definition of [4], we show that (1.4) is related to density for a random fields, and the random partition of [3]. An auxiliary computation is deferred to the appendix.

2 Random Field Construction

Let D be a closed convex subset of R^2 with smooth boundary. In R^2, choose coordinates (t, x) such that, for all $y \in D$, $t(y), x(y) > 0$. Let $\mathcal{L}_D$ denote the lines which intersect D, each line $\ell \in \mathcal{L}_D$ is parameterized by it's distance from the origin ρ_ℓ and the angle it forms with the $t = 0$ axis, α_ℓ.

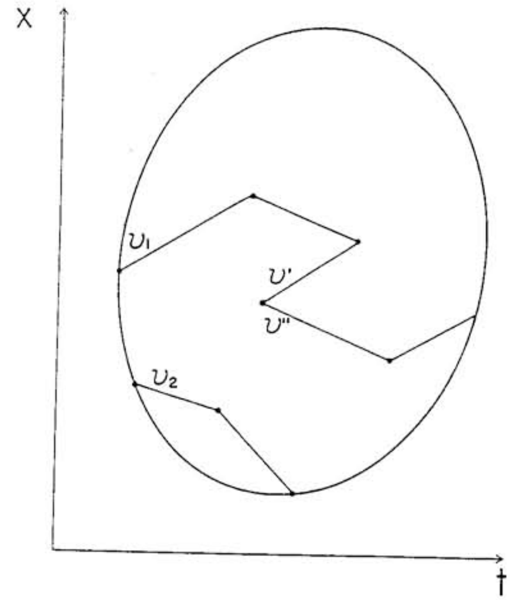

Figure 1: random polygonal partition

Let $\mu(d\ell)$ be a uniform measure on the set $\{\alpha_\ell, \rho_\ell \big| \ell \in \mathcal{L}_D\}$. The Poisson point process with intensity $\mu(d\ell)$ will be denoted μ^D, and the measure it induces on the boundary ∂D by the hitting points $\{(x_\ell, t_\ell)\big| t_\ell = \inf\{t \big| \ell \in D\}\}$ is again a Poisson point process on the triple (x, t, ν) with intensity $\mu^{\partial D}$. Here and in the sequel, ν denotes the velocity of the particle, i.e. the tangent of the angle formed by the trajectory and the t axis.

For any line $\ell \in \mathcal{L}_D$, let ν_ℓ denote the slope of ℓ. Clearly, $\mu(d\ell)$ can be considered as a measure on ν_ℓ and x_ℓ, the intersection of ℓ with the x axis. In the sequel, we consider the measure $\mu(d\nu, dx)$ obtained from the uniform measure $\mu(d\ell)$ on α, ρ, i.e.

$$\mu(d\ell) = \mu(d\nu,\ dx) = dx \frac{d\nu}{(1+\nu^2)^{1/2}} \tag{2.1}$$

Inside D, construct a point process on the quadruple $(t, y, \nu',\ \nu'')$ with intensity

$$\mu^P(dt,\ dy,\ d\nu',\ d\nu'') = |\nu' - \nu''| dy dt \frac{d\nu'}{(1+(\nu')^2)^{3/2}} \frac{d\nu''}{(1+(\nu'')^2)^{3/2}} \tag{2.2}$$

Finally, construct a random partition of D as follows:

Pick up on ∂D, n_0 triples (t, x, ν) according to the law $\mu^{\partial D}$, and inside D, n_1 quadruples (t, y, ν', ν'') according to the Poisson process with intensity μ^P. At each of those points, start a line of slope ν (two lines of slopes ν', ν'' in the case of interior points) and evolve

ν according to the Markov transition law

$$P(\nu_{t+dt} \in du \Big| \nu_t = \nu) = |u - \nu| \frac{dudt}{(1+u^2)^{3/2}} \tag{2.3}$$

Finally, at each intersection of lines (when viewing it in the direction of growing t) kill the intersected lines. Clearly, such dynamics describe a random partition of D by polygons, c.f. fig. 1.

The basic result of [3] is:

Lemma 2.1

$$P(n,\ \ell \in d\ell_i,\ i = 1, \cdots, n) = \frac{1}{n!} \mu(d\ell_1) \cdots \mu(d\ell_n) \exp(-2 \sum_{i=1}^{n} L(\ell_i)) \tag{2.4}$$

where $L(\ell_i)$ denotes the length of the i-th segment.

Note that due to the presence of n in (2.4), one can't consider (2.4) directly as a candidate for a density: indeed, if one were to consider $P(n,\ \ell_i \in \underline{\ell}_i \pm \epsilon)$, the required normalization constant (as $\epsilon \to 0$) would have depended on n and therefore, a path with no jumps will be infinitely more likely than a path with one jump.

One way out of this problem is by using an appropriate definition: Let

$$\begin{aligned} Z^n &\triangleq \int_{\mathcal{L}_D} \cdots \int_{\mathcal{L}_D} P(n, \ell_i \in d\ell_i,\ i = 1, \cdots,\ n) \\ Z &\triangleq \sum_{n=0}^{\infty} Z^n \end{aligned}$$

$(\frac{Z^n}{Z})$ is the probability of having n lines in a specific partition. Now, one may define:

Definition The prior density of a partition (n, ℓ_i) is given by

$$p(n, \ell_i) = (\frac{Z^n}{Z}) \lim_{\epsilon \to 0} \frac{P(n, \ell_i \in \ell_i \pm \epsilon,\ i = 1, \cdots, n)}{2\epsilon^{2n}} \tag{2.5}$$

We turn now to the random field part of the definition: given a partition $\Gamma_1, \cdots, \Gamma_n$, let z be the solution to the SPDE

$$\begin{aligned} \Delta z &= n \\ z\Big|_{\Gamma_i} &= 0 \end{aligned} \tag{2.6}$$

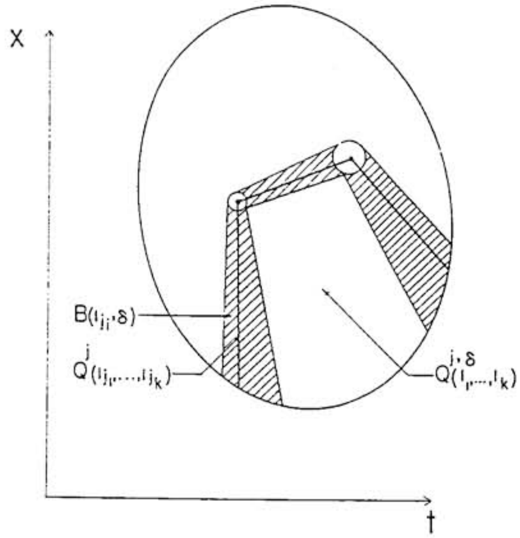

Figure 2: domain smoothing

Note that since the Γ_i define disjoint domains with Lipschitz boundaries, (2.6) define a set of disjoint, independent SPDE's which, in each domain, possess a unique weak solution, which is continuous up to the boundary (cf. [4] for details).

Let $Q^j(\ell_{j_1},\cdots,\ell_{j_k})$, $j<n+4$ denote the boundaries of one component of the partition (n,ℓ_i) where we allow for the boundaries of D to be counted as ℓ_i segments. Note that since, when changing partition boundaries, the set of candidates ϕ changes, we have to introduce a smoothing procedure on such candidates. Let therefore

$$Q^{j,\delta}(\ell_{j_1},\cdots,\ell_{j_k}) = \bigcap_{\ell_i\in B(\ell_{j_i},\delta)} Q^j(\ell_1,\cdots,\ell_k)$$

c.f. fig. 2. Let ϕ be given in the interior of $Q^j(\ell_{j_1}\cdots\ell_{j_k})$ with $\phi=0$ on $\ell_{j_1},\cdots,\ell_{j_k}$. Let $\hat{Q}^{j,\delta}(\ell_{j_1},\cdots,\ell_{j_k}) = \{x\in Q^{j,\delta}(\ell_{j_1},\cdots,\ell_{j_k})\big|B(x,\delta)\in Q^{j,\delta}(\ell_{j_1},\cdots,\ell_{j_k})\}$ denote a sub domain of $Q^{j,\delta}(\ell_{j_1},\cdots,\ell_{j_k})$. Note that for small enough δ,

$$d(\hat{Q}^{j,\delta}(\ell_j,\cdots,\ell_{j_k}),Q^j(\ell_1,\cdots,\ell_k)) \leq k_1\delta \tag{2.7}$$

where $d(A,B)$ denotes the Hausdorff distance between two sets and k_1 is a constant which depends only on the geometry of $Q^j(\ell_1,\cdots,\ell_k)$.

Finally, let $j^\delta(x) = C_\delta \exp(-1/(|x|^2 - (\frac{\delta}{2})^2))$ be a smooth mollifier of support $\frac{\delta}{2}$, let $\hat{\phi}$ denote the restriction of ϕ to $\hat{Q}^{j,\delta}(\ell_{j_1}, \cdots, \ell_{j_k})$, and let $\phi^\delta \triangleq \hat{\phi} * j^\delta$. Note that $\phi^\delta = 0$ on the boundary of $Q^{j,\delta}(\ell_{j_1}, \cdots, \ell_{j_k})$ (and therefore, necessarily, on the boundary of Q^j). The main technical lemma needed in the sequel is:

Lemma 2.2 Let $(\ell_{j_1}, \cdots, \ell_{j_k})$ be given, and let ϕ be given. Assume z satisfies

$$\begin{aligned} \Delta z &= n \\ z\Big|_{\hat{\ell}_{j_1}, \cdots, \hat{\ell}_{j_k}} &= 0 \end{aligned} \tag{2.8}$$

where $d(\ell_{j_i}, \hat{\ell}_{j_i}) < \delta$.

Then there exists a choice $\delta(\epsilon) \xrightarrow{\epsilon \to 0} 0$ such that

$$\log \frac{P(||z - \phi||_2 < \epsilon)}{P(||z||_2 < \epsilon)} \xrightarrow{\epsilon \to 0} -\frac{1}{2} \int_{Q(\ell_{j_1}, \cdots, \ell_{j_k})} |\Delta \phi|^2 \tag{2.9}$$

for all $\phi \in W_0^{4,2}(Q(\ell_{j_1}, \cdots, \ell_{j_k}))$, where $||\ ||_2$ denotes the norm in $L^2(Q(\ell_{j_1}, \cdots, \ell_{j_k}))$.

Proof The proof uses the machinery of [4]. We will emphasize here the new elements required to adapt it to our situation.

First, note that since $\phi \in W_0^{4,2}(Q(\ell_{j_1}, \cdots, \ell_{j_k}))$, one obtains immediately that for any choice $\delta(\epsilon) \xrightarrow{\epsilon \to 0} 0$, one has that

$$\int_{Q(\hat{\ell}_{j_1}, \cdots, \hat{\ell}_{j_k})} |\Delta \phi^\delta|^2 \xrightarrow{\epsilon \to 0} \int_{Q(\ell_{j_1}, \cdots, \ell_{j_k})} |\Delta \phi|^2 \tag{2.10}$$

Next, note that for δ small enough, $||\phi^\delta - \phi||_2 \le k(\delta)$ where $k(\delta) \underset{\delta \to 0}{\longrightarrow} 0$. Therefore, one obtains

$$P(||z - \phi^\delta||_2 < \epsilon - k(\delta)) \le P(||z - \phi||_2 < \epsilon) \le P(||z - \phi^\delta||_2 < \epsilon + k(\delta)) \tag{2.11}$$

As in [4 , eq. 3.3, 3.4], one has that

$$\frac{P(||z - \phi^\delta||_2 < \mu)}{P(||z||_2 < \mu)} = e^{-\frac{1}{2} \int_{Q(\hat{\ell}_{j_1}, \cdots, \hat{\ell}_{j_k})} |\Delta \phi^\delta|^2} \cdot E(\exp - \int_{Q(\hat{\ell}_{j_1}, \cdots, \hat{\ell}_{j_k})} (\Delta \phi^\delta) \circ n \Big| \, ||z||_2 < \mu) \tag{2.12}$$

Note however that

$$\begin{aligned} 1 &\le E(\exp - \int_{Q(\hat{\ell}_{j_1}, \cdots, \hat{\ell}_{j_k})} \Delta \phi^\delta \circ n \Big| \, ||z||_2 < \mu) \\ &= E(\exp - \int_{Q(\hat{\ell}_{j_1}, \cdots, \hat{\ell}_{j_k})} \Delta^2 \phi^\delta z \Big| \, ||z||_2 < \mu) \\ &\le \exp(\mu ||\Delta^2 \phi^\delta||_2) \underset{\mu \to 0}{\longrightarrow} 1 \end{aligned} \tag{2.13}$$

where we have used in the first equality the fact that ϕ^δ has compact support in $Q(\hat{\ell}_{j_1}, \cdots, \hat{\ell}_{jk})$. Substituting (2.13) in (2.12) and using (2.10) yields therefore, for any $\mu(\epsilon) \xrightarrow{\epsilon \to 0} 0$, $\delta(\epsilon) \xrightarrow{\epsilon \to 0} 0$:

$$\frac{P(||z - \phi^\delta||_2 < \mu(\epsilon))}{P(||z||_2 < \mu(\epsilon))} \underset{\delta(\epsilon) \to 0}{\overset{\mu(\epsilon) \to 0}{\longrightarrow}} \exp(-\frac{1}{2} \int_{Q(\ell_{j_1}, \cdots, \ell_{j_k})} |\Delta \phi|^2) \tag{2.14}$$

¿From (2.11) and (2.14) one concludes that, if one may show that there exist $\delta(\epsilon)$ such that

$$\frac{P(||z||_2 < \epsilon)}{P(||z||_2 < \epsilon + k(\delta(\epsilon)))} \xrightarrow{\epsilon \to 0} 1 \tag{2.15}$$

then the lemma holds. Note however that, if λ_i denotes the set of eigenvalues of (2.8), one has that

$$||z||_2 \overset{d}{=} (\sum_{i=1}^{\infty} \frac{x_i^2}{\lambda_i^2})^{1/2} \tag{2.16}$$

where x_i are independent standard normal random variables, and the sum in (2.16) is a.s. finite due to the fact that $\lambda_i \sim i$ as $i \to \infty$ (c.f. [1]) and that $\lambda_0 > 0$. Therefore, by choosing $\delta(\epsilon) \xrightarrow{\epsilon \to 0} 0$ fast enough, (2.15) holds. In the appendix, it is shown that taking $\delta(\epsilon) = \epsilon^3$ yields the required convergence.

□

Remark As in [4], $\phi \in W_0^{4,2}[Q(\hat{\ell}_1, \cdots, \hat{\ell}_k)]$ will not be quite enough for our needs for it will turn out that the optimizing path ϕ does not satisfy this condition. However, exactly as in [4], one may modify the proof above to include the solution to appropriate $\dot{y}$ driven SPDE's.

Combining lemmas 2.1 and 2.2, we are ready to state our

Theorem 2.1 *Let $(n, \Gamma_1, \cdots, \Gamma_n)$ denote a partition of D as described in the beginning of this section and let $\phi_1, \cdots, \phi_j$ denote functions on the appropriate sub domains of D which are zero on the partition segments ∂D and Γ_i. Assume that $\phi_i \in W_0^{4,2}(D_i)$ where D_i denotes the obvious subdomain associated with ϕ_i. Then*

$$\lim_{\epsilon \to 0} \frac{Z^n}{Z} \frac{P(n,\ \ell_i \in \Gamma_i \pm \epsilon,\ ||z - \phi||_2 < \epsilon)}{P(||z||_2 < \epsilon) 2\epsilon^{2n}} = \frac{Z^n}{Z} e^{-2\ell(\Gamma)} e^{-\frac{1}{2}\int_{D \setminus (\Gamma \cup \partial D)} |\Delta \phi|^2} \tag{2.17}$$

Remarks It seems appropriate, at this point, to compare (2.17) and (1.3) with $\alpha, \beta \to \infty$ and $\alpha/\beta = 1$. The differences are apparent: first, in (2.17) there is a penalty on the total number of partition lines in terms of Z^n. This is reminiscent of Richardson's [6] constraint on the "number of components", although it arises here from entirely different reason and has a completely different interpretation. By a choice of different neighborhoods in (2.5) (in particular, n dependent neighborhoods), one can get rid of this term.

Next, (2.17) is defined only over straight line boundaries, thus restricting significantly the scope of the optimization over (1.3).

Finally, note that in (2.17) the $|\nabla\phi|^2$ term has been replaced by a $|\Delta\phi|^2$ term. The reason for that can be traced to the model (2.8): indeed, had we tried to write instead of $\Delta z = n$ an equation of the form $\Delta^{1/2}\hat{z} = n$, only distributional solutions would have existed for $\hat{z}$. One could try then to use, instead of L_2 neighborhoods, distributional neighborhoods of $\hat{z}$. We do not follow that approach here, since it seems unnatural in this context.

We may obtain, using [4, lemma 4.1], the following conditional form of theorem 2.1:

Corollary 2.1 Let $(n, \Gamma_1, \cdots, \Gamma_n)$ and $(\phi_1, \cdots, \phi_n)$ be as in theorem 2.1, and let $\mathcal{F}_y$ denote the sigma field generated by the observations $y_{\cdot}$. Then,

$$\lim_{\epsilon\to 0} \frac{Z^n}{Z} \frac{P(\ell_i \in \Gamma_i \pm \epsilon,\ ||z-\phi||_2 < \epsilon \big| \mathcal{F}_y)}{P(||z||_2 < \epsilon) 2\epsilon^{2n}} = k(y) \frac{Z^n}{Z} e^{-2\ell(\Gamma)} e^{-\frac{1}{2}\int_{D\setminus(\Gamma\cup\partial D)} |\Delta\phi|^2} e^{(\int_D \phi\circ\dot{y} - \frac{1}{2}\int_D \phi^2)} \tag{2.18}$$

where $k(y)$ denotes a $\mathcal{F}_y$ measurable random variable which does not depend on the specific ϕ chosen.

It is worthwhile to remark that we do not have at this point a satisfying existence theorem for (2.18): indeed, unlike the situation in [4], the fact that the boundaries change on a minimizing sequence may prevent the existence of a minimizing function in $W_0^{2,2}(D_i)$. On the other hand, working in $W^{2,2}(D_i)$ isn't helpful because, unlike in the gradient case treated in [6], the Laplacian isn't coercive over $W^{2,2}(D_i)$.

Appendix

In this appendix, we give a way to compute explicitly the exact $\delta(\epsilon)$ needed in the proof of lemma 2.2. Indeed, from (2.12) finding the appropriate $\delta(\epsilon)$ entails finding the asymptotics of $P(y < \epsilon)$ as $\epsilon \to 0$, where $y \triangleq \sum_{i=1}^{\infty} \frac{x_i^2}{\lambda_i^2}$, x_i are a sequence of independent, identically distributed Normal random variables, and $\sum_{i=1}^{\infty} \frac{1}{\lambda_i^2} < \infty$. We note that the bounds we obtain, namely (3.2) and (3.6) below, are tighter than the respective (more general) bounds in [7], equations (4.5.1) and (4.5.2).

To compute the required asymptotics, we make use of large deviations techniques. Since y doesn't possess exponential moments, a direct attack via Cramer's theorem seems impossible. On the other hand, note that, by Chebycheff's inequality, using the fact that $y \geq 0$,

$$P(\frac{y}{\epsilon} < 1) \leq E(e^{-s(\frac{y}{\epsilon}-1)}) = e^{s} \cdot \frac{1}{\prod_{i=1}^{\infty}(1+\frac{2s}{\epsilon\lambda_i^2})^{1/2}} \tag{3.1}$$

for all $s > 0$. In particular, one obtains that

$$\log P(\frac{y}{\epsilon} < 1) \leq s - \frac{1}{2}\sum_{i=1}^{\infty}\log(1+\frac{2s}{\epsilon\lambda_i^2}) \tag{3.2}$$

On the other hand, let

$$d\nu^{s}(z) \triangleq \frac{e^{-sz}d\mu(\epsilon z)}{\int_0^{\infty} e^{-sz}d\mu(\epsilon z)}, \quad s \geq 0 \tag{3.3}$$

where μ denotes the measure defined by the random variable y. Let s_ϵ be defined by the relation

$$1 = \int z d\nu^{s_\epsilon}(z) \tag{3.4}$$

which leads to the equation

$$1 = \frac{2}{\epsilon}\sum_{i=1}^{\infty}\left(\frac{1/\lambda_i^2}{1+\frac{2s_\epsilon}{\epsilon\lambda_i^2}}\right) \tag{3.5}$$

By the standard change of measure argument, noting that $\nu^{s_\epsilon}(B(1,\eta)) \underset{\epsilon\to 0}{\longrightarrow} 1$ for any $\eta > 0$ by the law of large numbers, one obtains that, for ϵ small enough,

$$\log P(\frac{y}{\epsilon} < 1) \geq s_\epsilon - \frac{1}{2}\sum_{i=1}^{\infty}\log\left(1+\frac{2s_\epsilon}{\epsilon\lambda_i^2}\right) + o(1) \tag{3.6}$$

Since, as $i \to \infty$, $\lambda_i \sim i$, one obtains that $s_\epsilon \sim \frac{c}{\epsilon}$ for small enough ϵ and therefore, comparing (3.6) and (3.2), one obtains that

$$\log P(\frac{y}{\epsilon} < 1) \sim -\frac{1}{2}\sum_{i=1}^{\infty}\log(1+\frac{2s_\epsilon}{\epsilon\lambda_i^2})$$

which implies

$$\log P(\frac{y}{\epsilon} < 1) \sim (\log\epsilon^2)\epsilon^{-1}$$

To find an appropriate $\delta(\epsilon)$, it is therefore enough to have

$$(\epsilon+\delta(\epsilon))^{-1}\log(\epsilon+\delta(\epsilon)) - \epsilon^{-1}\log(\epsilon) \underset{\epsilon\to 0}{\longrightarrow} 0$$

which is clearly satisfied with $\delta(\epsilon) = \epsilon^3$.

□

Acknowledgments This work was supported by the Army Research Office under contract DAAL03-86-K-0171. The work of the second author was done while he visited LIDS at MIT.

References

[1] Agmon, S., Lectures on elliptic boundary value problems, Van Nostrand, 1965.

[2] Ambrosio, L., Existence theory for a new class of variational problems, preprint (1988), to appear.

[3] Arak, T. and Surgailis, D., Markov fields with polygonal realizations, Prob. Th. Rel. Fields, 80 (1989), pp. 543-579.

[4] Dembo, A. and Zeitouni, O., Maximum a-Posteriori estimation of elliptic Gaussian fields observed via a noisy nonlinear channel, submitted.

[5] Mumford, D. and Shah, J., Optimal approximations by piecewise smooth functions and associated variational problems, Comm. Pure and App. Math, XLII (1989), pp. 577-685.

[6] Kulkarni, S.R., Mitter, S.K., and Richardson, T., An existence result and lattice approximation for a variational problem arising in computer vision, to appear, Proceedings of the Signal Proc. Workshop, Inst. for Math. and its Appl. Also, Report CICS-P-123, MIT, 1989.

[7] Hoffman-Jørgensen, J., Shepp, L.A., and Dudley, R.M., On the lower tails of Gaussian seminorms, The Annals of Probability, 7 (1979), pp. 319-342.