A canonical process for estimation of convex functions: the "invelope" of integrated Brownian motion $+t^4$

By Piet Groeneboom, Geurt Jongbloed, and Jon A. Wellner[2]

Delft University of Technology and Vrije Universiteit, Vrije Universiteit, and University of Washington

A process associated with integrated Brownian motion is introduced that characterizes the limit behavior of nonparametric least squares and maximum likelihood estimators of convex functions and convex densities, respectively. We call this process "the invelope" and show that it is an almost surely uniquely defined function of integrated Brownian motion. Its role is comparable to the role of the greatest convex minorant of Brownian motion + a parabolic drift in the problem of estimating monotone functions. An iterative cubic spline algorithm is introduced that solves the constrained least squares problem in the limit situation and some results, obtained by applying this algorithm, are shown to illustrate the theory.

1. Introduction. Consider the following nonparametric estimation problem: $X_1, \ldots, X_n$ is a sample of observations, generated by a density f with the property that $f^{(k)}$ is monotone on the support of the distribution of the X_i, where k

[1]Received

AMS 2000 subject classifications. Primary: 62G05; secondary 60G15, 62E20.

Key words and phrases. convex function,estimation,Gaussian process, integrated Brownian motion,least squares

[2] Research supported in part by National Science Foundation grant DMS-95-32039, NIAID grant 2R01 AI291968-04, and the Stieltjes Institute

is fixed and ≥ 0. (Here and in the following, for a real valued function g defined on some subset of the real line, $g^{(k)}$ denotes the kth derivative of g. We also use the usual prime notation $g' = g^{(1)}$ for the first derivative, $g'' = g^{(2)}$ for the second derivative, and $g^{(0)}$ is simply the function g itself.) A well-known example of this situation is when $k = 0$: then f is a decreasing density on $[0, \infty)$. In that case there is a well-known nonparametric maximum likelihood estimator: the Grenander estimator, that is defined as the left-continuous slope of the least concave majorant of the empirical distribution function of the X_i's. The asymptotic behavior of the Grenander estimator, the (nonparametric) maximum likelihood estimator of f, is well studied, and it is known (for example) that, if $\hat{f}_n$ denotes the Grenander estimator, and if f has a strictly negative derivative $f'(t_0)$ at $t_0 \in (0, \infty)$, that

$$\frac{n^{1/3}\left\{\hat{f}_n(t_0) - f(t_0)\right\}}{\left\{\frac{1}{2}f(t_0)|f'(t_0)|\right\}^{1/3}} \xrightarrow{\mathcal{D}} 2Z, \tag{1.1}$$

where $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution, and $2Z$ is the slope of the (greatest) convex minorant of $\{W(t) + t^2 : t \in \mathbb{R}\}$ at zero, where W is two-sided Brownian motion, originating from zero; see, e.g., Prakasa Rao (1969). An alternative interpretation of the limit distribution is that Z is the location of the minimum of $\{W(t) + t^2 : t \in \mathbb{R}\}$, where W is again two-sided Brownian motion, originating from zero; see Groeneboom (1985), and, for computation of the distribution of Z, see Groeneboom (1988) and Groeneboom and Wellner (2000).

But now consider, for example, the estimation problem in the situation where we assume that f' is increasing ($k = 1$), and f is decreasing on $[0, \infty)$, so f is a convex decreasing density on $[0, \infty)$. In this case, a result of type (1.1) is not known, and there are only partial results, telling us, for example, that for fixed $t_0 \in (0, \infty)$,

$$n^{2/5}\left\{\hat{f}_n(t_0) - f(t_0)\right\} = \mathcal{O}_p(1), \tag{1.2}$$

see, e.g., JONGBLOED (1995).

Similarly, let Y_i, $i = 1, \ldots, n$ be observations in a *regression* setting:

$$Y_i = \theta\left(t_{n,i}\right) + e_i,\ i = 1, \ldots, n,\ t_{n,i} = i/n,$$

where the e_i are i.i.d. random variables with expectation zero and finite variance $\sigma^2 > 0$. In this situation one can consider the problem of estimating the regression function θ under the restriction that $\theta^{(k)}$ is monotone for some $k \geq 0$. For this situation Theorem 5.2 in BRUNK (1970) tells us that, if θ is monotone ($k = 0$), the isotonic least squares estimator $\hat{\theta}_n$ of the function θ has the property

$$\frac{n^{1/3}\left\{\hat{\theta}_n(t_0) - \theta(t_0)\right\}}{\left\{\frac{1}{2}\theta'(t_0)\sigma\right\}^{1/3}} \xrightarrow{\mathcal{D}} Z, \tag{1.3}$$

at a fixed point $t_0 \in (0, 1)$, where Z is the slope of the (greatest) convex minorant of $\{W(t) + t^2 : t \in \mathbb{R}\}$ at zero and where it is assumed that θ has a continuous derivative $\theta'(u) \neq 0$ in a neighborhood of t_0. Here W is, as before, two-sided Brownian motion, originating from zero.

We now can again consider the estimation problem in the situation where we assume that θ' is increasing ($k = 1$), so θ is a *convex* regression function on $[0, 1]$. In this case, a result of type (1.3) is not known, and there are again only partial results, telling us for example that

$$n^{2/5}\left\{\hat{\theta}_n(t_0) - \theta(t_0)\right\} = \mathcal{O}_p(1), \tag{1.4}$$

where $\hat{\theta}_n$ is the least-squares estimator of θ; see, e.g., MAMMEN (1991).

In WANG (1994) it is stated that in this situation we have, at a point $t_0 \in (0, 1)$, under the additional conditions that $E\exp(ue_i^2) < \infty$, for some $u > 0$ and that $\theta''(t_0)$ exists and is strictly positive,

$$n^{2/5}\left\{\frac{6}{\theta''(t_0)\sigma^4}\right\}^{1/5}\left(\hat{\theta}_n(t_0) - \theta(t_0)\right) \xrightarrow{\mathcal{D}} F, \tag{1.5}$$

where F is the limiting distribution of $f_c(0)$, as $c \to \infty$, and where f_c is the minimizer of

$$\int_{-c}^{c} f(t)^2\,dt - 2\int_{-c}^{c} f(t)\,d\bigl(W(t)+t^3\bigr) \tag{1.6}$$

over the class of convex functions on $(-c,c)$, under a boundary restriction on the values of $f(-c)$ and $f(c)$. Actually, in WANG (1994) concave instead of convex functions are considered, but this is essentially the same problem, and we only changed some signs to change the statement into a statement on the estimation of convex functions.

The following heuristic argument makes this statement "easy to believe". Assume for simplicity (and in fact, without loss of generality) that $\theta(t_0)=0$ and $\theta'(t_0)=0$. Let $\hat\theta_n$ be the least squares estimator of the convex function θ. It then follows from MAMMEN (1991) that $\hat\theta_n$ is a piecewise linear function with changes of slope at a distance of order $n^{-1/5}$ in a neighborhood of t_0 and that, on an interval $J_n = [t_0 - cn^{-1/5}, t_0 + cn^{-1/5}]$, with $c>0$, we have the relation

$$\begin{aligned}&\sum_{t_{n,i}\in J_n} \hat\theta_n(t_{n,i})^2 = n \sum_{t_{n,i}\in J_n} \hat\theta_n(t_{n,i})^2 \left\{t_{n,i}-t_{n,i-1}\right\}\\ &\sim n\int_{t_0-cn^{-1/5}}^{t_0+cn^{-1/5}} \hat\theta_n(t)^2\,dt = \int_{-c}^{c} n^{4/5}\hat\theta_n(t_0+n^{-1/5}t)^2\,dt = \int_{-c}^{c} \hat f_n(t)^2 dt,\end{aligned}$$

where $\hat f_n$ is the obvious rescaling of the convex function $\hat\theta_n$:

$$\hat f_n(t) = n^{2/5}\hat\theta_n\left(t_0+n^{-1/5}t\right) = n^{2/5}\left\{\hat\theta_n\left(t_0+n^{-1/5}t\right)-\theta(t_0)\right\},\ t\in[-c,c].$$

Using the same rescaling, we can write

$$\begin{aligned}\sum_{t_{n,i}\in J_n} \hat\theta_n(t_{n,i})Y_i &\sim \sum_{t_{n,i}\in J_n} \hat\theta_n(t_{n,i})\left\{Y_i - \theta(t_{n,i}) + \frac12\theta''(t_0)(t_{n,i}-t_0)^2\right\}\\ &= \sum_{t_{n,i}\in J_n} \hat\theta_n(t_{n,i})\left\{e_i + \frac12\theta''(t_0)(t_{n,i}-t_0)^2\right\}\\ &\sim \int_{-c}^{c} \hat f_n(t)\,d\left\{\sigma W(t) + \frac16\theta''(t_0)t^3\right\},\end{aligned}$$

where $\sum_{t_{n,i}\in J_n}\hat\theta_n(t_{n,i})e_i=\sum_{t_{n,i}\in J_n}\hat f_n\left(n^{1/5}(t_{n,i}-t_0)\right)n^{-2/5}e_i$ and W is standard two-sided Brownian motion, originating from zero. Hence,

$$\sum_{t_{n,i}\in J_n}\{(\hat\theta_n(t_{n,i})-Y_i)^2-Y_i^2\}\sim\int_{-c}^{c}\hat f_n(t)^2\,dt-2\int_{-c}^{c}\hat f_n(t)\,d\left\{\sigma W(t)+\frac{1}{6}\theta''(t_0)t^3\right\},$$

where $\hat\theta_n$ minimizes

$$\sum_{t_{n,i}\in J_n'}\{\theta_n(t_{n,i})-Y_i\}^2\ \text{ and therefore also }\sum_{t_{n,i}\in J_n'}\left\{(\theta_n(t_{n,i})-Y_i)^2-Y_i^2\right\}$$

for convex functions θ_n on intervals $J_n'\supset J_n$, having as endpoints locations of changes of slope of $\hat\theta_n$. This makes it plausible that the linearly to $\mathbb{R}$ extended function $\hat f_n$ converges in distribution, in the topology of uniform convergence on compacta, to the limit of the functions f_c as $c\to\infty$, minimizing

$$\int_{-c}^{c}f(t)^2\,dt-2\int_{-c}^{c}f(t)\,d\left\{\sigma W(t)+\frac{1}{6}\theta''(t_0)t^3\right\},$$

over convex functions f on $[-c,c]$, under certain boundary conditions at $-c$ and c (the influence of which will die out in bounded intervals, as $c\to\infty$), provided such a limit exists. By Brownian scaling arguments (see section 5) this would be equivalent to saying that the rescaled functions $\tau\mapsto a\hat f_n(ta^2\sigma^2)$ with $a=(6/\theta''(t_0)\sigma^4)^{1/5}$ converge in distribution to the limit of the functions $\tilde f_c$, as $c\to\infty$, minimizing (1.6). This would in particular mean that (1.5) holds, provided $\lim_{c\to\infty}\tilde f_c(0)$ exists.

However, the proof of this “easy to believe” statement in Wang (1994) contained several flaws. For example, in proving that the value of $\hat f_n(0)$ stabilizes, as $n\to\infty$, it was assumed that the changes of slope of $\hat f_n$ in a finite interval $[-c,c]$ are all bigger than $\theta''(t_0)/2$, for large n, by mistakenly assuming that the constrained regression problem can be solved by considering, at a finite number of points, *separately* regression on the deterministic function θ and regression on the noise variables e_i. Then, since the (constrained) regression on the (“true”) deterministic function would lead to a piecewise linear function, having changes of slope bigger than $\theta''(t_0)+o_p(1)$, and the (constrained) regression on the errors e_i would lead to an almost constant function, one would get that the changes of slope of $\hat{f}_n$ in a finite interval $[-c,c]$ are all bigger than $\theta''(t_0)/2$ for large n. But one clearly cannot split the constrained regression problem in this way.

There is no a priori reason to assume that the changes of slope of $\hat{f}_n$ in a finite interval $[-c,c]$ are all bigger than $\theta''(t_0)/2$ for large n, and we think that this assumption is false, both for the finite sample solution $\hat{f}_n$ and for the functions f_c, used in the limit situation. Moreover, in comparing two solutions f_c and $\tilde{f}_c$ with different boundary conditions at $-c$ and c, with the aim of showing that the influence of the boundary conditions "dies out" as $c \to \infty$, only functions with the same locations of changes of slope were compared in Wang (1994) (in the finite sample situation), whereas different boundary conditions will generally lead to different locations of changes of slope of the functions f_c and $\tilde{f}_c$ (see section 3). In this sense the situation is strikingly different from the situation for the estimation of *monotone* functions, where the set of locations of jumps of a constrained solution on an interval $[-c,c]$ will be a subset of the set of locations of change of slope of the greatest convex minorant of $\{W(t)+t^2 : t \in I\!\!R\}$!

In fact, up till now, it has not even been proved that a function f_c, minimizing (1.6), under, say, the boundary conditions $f(c) = f(-c) = 3c^2$, has *isolated* points of change of slope. If all changes of slope were bigger than a fixed constant, as assumed in Wang (1994), this would be automatically fulfilled. But since we cannot make that assumption, we can also not assume that the points of change of slope are isolated.

We have described the difficulties of the approach in Wang (1994) in some detail in an attempt to explain why the problem of characterizing the limit distribution has been open for so many years and also to give an idea of the difficulties involved here. But we are of course indebted to Wang (1994) for putting us on the track of proving that the limit distribution is given by the limit of the function f_c, defined in (1.6), which we want to gratefully acknowledge here.

The problems with the arguments in Wang (1994) led us to to try a whole new "geometrical" approach to this problem. In the estimation problem for *monotone* functions, the limit behavior is described by a "canonical" function of the process $\{W(t)+t^2 : t \in \mathbb{R}\}$: its greatest convex minorant. Let X be the process $\{X(t) : t \in \mathbb{R}\} = \{W(t)+t^2 : t \in \mathbb{R}\}$ and let C be its greatest convex minorant. Then it is not hard to show that the slope of the greatest convex minorant C of X at a 0 is the limit of $f_c(0)$, where $c \to \infty$, and f_c minimizes

$$\int_{-c}^{c} f(t)^2\,dt - 2\int_{-c}^{c} f(t)\,d\big(W(t)+t^2\big) \tag{1.7}$$

over all *nondecreasing* functions $f : [-c,c] \to \mathbb{R}$, under the boundary constraints $f(-c) = -2c$, $f(c) = 2c$. In this case the proofs are relatively easy, since we know, for example from the jump process characterization in Groeneboom (1988), that the points of jump of the slope of the greatest convex minorant *are* indeed isolated (although the size of a jump can be arbitrarily small!) and since the constrained minimization problem also has a solution in terms of a greatest convex minorant function. But all these arguments really rely on the explicit characterization in terms of the greatest convex minorant and we do not have something similar for the estimation problem in the case of *convex* functions. So this motivates the search for a "canonical" process that, for the estimation of convex functions, plays a role similar to the role of the greatest convex minorant in the estimation of monotone functions.

We found such a canonical process for the estimation problem of convex functions and we coined the term "invelope" for it (motivated by the terminology "convex envelope" in the estimation problem of monotone functions). It is a twice continuously differentiable function H with a *convex* second derivative and the property that $H \geq Y$ (so the graph of H lies *inside* the graph of Y), where Y is the process

$$\{Y(t) : Y(t) = V(t) + t^4,\, t \in \mathbb{R}\},$$

and where V is *integrated* Brownian motion, originating from zero.

The full characterization of the "invelope" is given in Theorem 1 in section 2. This is an almost surely uniquely defined function of integrated Brownian motion and its properties can be used to show that indeed $f_c(0)$, where f_c is the minimizer of (1.6) under the boundary conditions $f(-c) = k_1(c)$ and $f(c) = k_2(c)$, where $k_1(-c) - 3c^2$ and $k_2(c) - 3c^2$ are uniformly bounded as functions of c, converges almost surely to a finite limit, as $c \to \infty$. For convenience we changed $W(t) + t^3$ to $W(t) + 4t^3$, since the really important object is $V(t) + t^4$, where V is *integrated* Brownian motion, and therefore our boundary condition is that $k_1(-c) - 12c^2$ and $k_2(c) - 12c^2$ are uniformly bounded, but this makes no difference for the argument. In fact $f_c(0)$ converges almost surely to the *second derivative* of the "invelope" H at zero, as $c \to \infty$, see Corollary 4 in section 2. Corollary 4 also shows that indeed the influence of the boundary conditions dies out on fixed intervals, as $c \to \infty$, see the remark following this corollary.

However, proving that an object like our "invelope" indeed exists and is an (almost surely) uniquely defined function of integrated Brownian motion was the real bottleneck in getting any asymptotic distribution theory for the estimators in the convex estimation problem going. We believe that we have taken that hurdle in the present manuscript. The asymptotic distribution theory for the convex density and regression problems is treated in the companion paper to the present paper, Groeneboom, Jongbloed and Wellner (2001a).

We also hope that our treatment of the convex case opens the way for the treatment of the general estimation problem of a function f, under the restriction that $f^{(k)}$ is monotone, for some $k \geq 0$ (where one will have to study k times integrated Brownian motion).

2. The Gaussian problem: characterization of the solution Let $X(t) = W(t) + 4t^3$ where $W(t)$ is standard two-sided Brownian motion starting from 0, and define

$$Y(t) = \begin{cases} \int_0^t W(s)ds + t^4, & t \geq 0 \\ \int_t^0 W(s)ds + t^4, & t \leq 0\,. \end{cases} \tag{2.1}$$

Our main goal in this section is to prove the following theorem.

Theorem 1. *There exists an almost surely uniquely defined random continuous function H satisfying the following conditions:*

(i) The function H is everywhere above Y:

$$H(t) \geq Y(t), \text{ for each } t \in \mathbb{R}\,. \tag{2.2}$$

(ii) H has a convex second derivative.

(iii) H satisfies

$$\int_{\mathbb{R}} \{H(t) - Y(t)\}\, dH^{(3)}(t) = 0. \tag{2.3}$$

Note that condition (iii), in the presence of (i), means that the (increasing) function $H^{(3)}$ cannot change (i.e. increase) in a region where (i) is satisfied with strict inequality. The analogue in the monotone situation is that the slope of the convex minorant of the drifting Brownian motion cannot change at points where this minorant is strictly smaller than the drifting Brownian motion.

In particular, the probability that the convex function $H^{(2)}$ will have a change of slope at zero is equal to zero, meaning that the third derivative $H^{(3)}$ is almost surely well-defined at zero, see Corollary 1.

To prove Theorem 1, we first consider convex functions f_c, defined on intervals $[-c, c]$, that are approximations to the second derivative of our "invelope" on these intervals. Let the functional $\phi_c(g)$ be defined by

$$\phi_c(g) = \tfrac{1}{2} \int_{-c}^{c} g^2(t)\, dt - \int_{-c}^{c} g(t)\, dX(t), \tag{2.4}$$

for convex functions $g : [-c, c] \mapsto I\!R$. Consider the problem of minimizing $\phi_c(g)$ under the side constraints

$$g(-c) = k_1,\ g(c) = k_2, \tag{2.5}$$

and let the (allowed) set of convex functions g be defined by

$$\mathcal{G}(c, k_1, k_2) \equiv \{g : [-c, c] \to I\!R,\ g \text{ is convex},\ g(-c) = k_1,\ g(c) = k_2\}. \tag{2.6}$$

Then we have the following lemma.

Lemma 1. *For each fixed $c > 0$ and $k_1, k_2 \in I\!R$, the problem of minimizing $\phi_c(g)$ over $\mathcal{G}(c, k_1, k_2)$ has a unique solution $f \equiv f_{c,k_1,k_2}$.*

Proof: It is easily seen that a minimizer of (2.4) subject to (2.5) must be in a compact subset

$$\mathcal{G}(c, M, k_1, k_2) \equiv \{g \in \mathcal{G}(c, k_1, k_2) : \quad g(t) \geq -M \text{ for all } t \in [-c, c]\}$$

for some $0 < M < \infty$. To see this, note that if there is some $t_0 \in (-c, c)$ such that $g(t_0) < -M$, $|g(t)| > M/2$ on an interval of strictly positive length (nonvanishing as $M \to \infty$). This means that the first term in $\phi_c(g)$ is of order M^2 and the second of order M as $M \to \infty$. Comparing this to the value of ϕ_c attained at the linear function g_0 which joins $(-c, k_1)$ to (c, k_2), the claim follows.

Then existence follows from compactness of $\mathcal{G}(c, M, k_1, k_2)$ in e.g. the uniform topology together with continuity of ϕ_c on this set. Uniqueness follows from the strict convexity of ϕ_c and convexity of $\mathcal{G}(c, k_1, k_2)$: for $\lambda \in (0,1)$ and $f, g \in \mathcal{G}(c, k_1, k_2)$ we have

$$\phi_c(\lambda f + (1-\lambda)g) - \lambda\phi_c(f) - (1-\lambda)\phi_c(g) = -\frac{\lambda(1-\lambda)}{2}\int_{-c}^{c}\{f(t) - g(t)\}^2\,dt < 0$$

if $\int_{-c}^{c}\{f(t) - g(t)\}^2\,dt > 0$, and thus ϕ_c is strictly convex. □

For a fixed point t, the probability that Y will have a one-sided parabolic tangent at t, in the sense that there exists a second degree polynomial P such that $P(t) = Y(t)$, $P'(t) = Y'(t) = X(t)$ and $P(u) \geq Y(u)$ (or $P(u) \leq Y(u)$) for u in a neighborhood of t, is zero since Brownian motion is of infinite variation. For this reason we will assume in the following that $-c$ and c are points where such a one-sided derivative of Y does *not* exist.

The following characterization of the solution f_{c,k_1,k_2} of the minimization problem, considered in Lemma 1, will play a crucial role in our further development.

LEMMA 2. *(Characterization of the solution on a finite interval) Suppose that f is a convex function on $[-c, c]$ with second integral H, satisfying $H(-c) = Y(-c)$ and $H(c) = Y(c)$, i.e. $H'' = f$ and H is determined by its two values at $-c$ and c. Furthermore, suppose that Y does not have parabolic tangents at $-c$ and c. Then f minimizes $\phi_c(g)$ over $\mathcal{G}(c, k_1, k_2)$ if and only if the following conditions are satisfied:*

(2.7) $$H(t) \geq Y(t),\ t \in [-c, c],$$

$$\int_{(-c,c)} \{H(t) - Y(t)\}\, df'(t) = 0, \tag{2.8}$$

and

$$f(-c) = k_1, \quad f(c) = k_2. \tag{2.9}$$

Proof: Fix ω such that the parabolic tangents as described above, do not exist at $\pm c$. Suppose that H, F and f satisfy the conditions of the lemma where F is the derivative of H and f is the derivative of F. Let f' be (a version of) the derivative of f. Furthermore, let λ_1 and λ_2 be defined by

$$\lambda_1 = F(-c) - X(-c),\, \lambda_2 = X(c) - F(c), \tag{2.10}$$

and let the extended criterion function $\phi_{c,\lambda}$ be defined by

$$\phi_{c,\lambda}(g) = \phi_c(g) + \lambda_1\{g(-c) - k_1\} + \lambda_2\{g(c) - k_2\},$$

where $\lambda = (\lambda_1, \lambda_2)'$. Then, since

$$g^2 - f^2 = (g-f)^2 + 2f(g-f) \geq 2f(g-f), \tag{2.11}$$

we get for any convex function $g : [-c, c] \mapsto I\!R$,

$$\begin{aligned}\phi_{c,\lambda}(g) - \phi_{c,\lambda}(f) \geq& \int_{-c}^{c} f(t)\{g(t) - f(t)\}\, dt - \int_{-c}^{c} \{g(t) - f(t)\}\, dX(t)\\ &+\lambda_1\{g(-c) - f(-c)\} + \lambda_2\{g(c) - f(c)\}.\end{aligned}$$

Suppose (as we may) that the derivative g' of g has finite limits at $-c$ and c. Then integration by parts yields, using (2.10) and (2.8),

$$\begin{aligned}&\int_{-c}^{c} f(t)\{g(t) - f(t)\}\, dt - \int_{-c}^{c} \{g(t) - f(t)\}\, dX(t)\\ &\qquad +\lambda_1\{g(-c) - f(-c)\} + \lambda_2\{g(c) - f(c)\}\\ &= \int_{-c}^{c} \{X(t) - F(t)\}\{g'(t) - f'(t)\}\, dt = \int_{-c}^{c} \{X(t) - F(t)\}\, g'(t)\, dt\\ &= \int_{(-c,c)} \{H(t) - Y(t)\}\, dg'(t).\end{aligned} \tag{2.12}$$

In the last equality we use that the derivative f' of f has finite limits at $-c$ and c. This follows from the fact that we assume that $-c$ and c are points where Y does not have a one-sided parabolic tangent, implying that $F(-c) = X(-c)$ or $F(c) = X(c)$ cannot occur. Since $H \geq Y$, this implies $F(-c) > X(-c)$ and $F(c) < X(c)$. For $F(c) > X(c)$ would imply that $H(x) < Y(x)$ in a left neighborhood of c, since $H(c) = Y(c)$, and this contradicts $H \geq Y$. Similarly, $F(-c) < X(-c)$ cannot occur.

This, in turn, implies that f' has a finite limit at $-c$ and c. For since $F(c) < X(c)$, there exists a left neighborhood $(c-\delta, c)$ of c such that $F(t) < X(t)$, if $t \in (c-\delta, c)$. In a similar way there exists a right neighborhood $(-c, -c+\delta')$ of $-c$ such that $F(t) > X(t)$ for $t \in (-c, -c+\delta')$. Using that $H(t) > Y(t)$ for all t in a left (reduced) neighborhood of c, so that f behaves linearly on this set, we get the following implication. If $f'(t) \to \infty$, as $t \uparrow c$, then

$$\int_{c-\delta}^{u} f'(t)\{X(t) - F(t)\}\, dt \to \infty, \ \text{ as } u \uparrow c. \tag{2.13}$$

Similarly we would get

$$\int_{u}^{-c+\delta} f'(t)\{X(t) - F(t)\}\, dt \to \infty, \ \text{ as } u \downarrow -c, \tag{2.14}$$

if $f'(t) \to -\infty$, as $t \downarrow -c$. But since

$$\begin{aligned} &\int_{-c}^{c} f'(t)\,\{X(t) - F(t)\}\ dt \\ &= k_2\{X(c) - F(c)\} - k_1\{X(-c) - F(-c)\} + \int_{-c}^{c} f(t)^2\, dt - \int_{-c}^{c} f(t)\, dX(t) \end{aligned}$$

is finite, neither of these possibilities can occur. Note that (2.14) tends to ∞, if $f'(t) \to -\infty$, as $t \downarrow -c$, so we are not in a situation where positive infinite growth at c could be compensated by a piece of the integral tending to $-\infty$ as $u \downarrow -c$.

Now, if g is a function of the following type:

$$g(t) = a + bt + \sum_{i=1}^{k} a_i (t - t_i)_+, \tag{2.15}$$

where $-c < t_1 < \ldots < t_k < c$, $a, b \in I\!\!R$ and $a_i > 0$, for each $i = 1, \ldots, k$, we get

$$\int_{(-c,c)} \{H(t) - Y(t)\}\, dg'(t) = \sum_{i=1}^{k} a_i \{H(t_i) - Y(t_i)\} \geq 0,$$

using $H \geq Y$. Hence it follows that

$$\phi_{c,\lambda}(g) \geq \phi_{c,\lambda}(f) \tag{2.16}$$

for all g of the form (2.15). Now for an arbitrary convex function g on $[-c, c]$ there exists a sequence of functions $\{g_k\}$ of the form (2.15) with $\|g_k - g\|_\infty = \sup_{|t| \leq c} |g_k(t) - g(t)| \to 0$ as $k \to \infty$ (where $\|\cdot\|_\infty$ is the uniform norm). It follows from the continuity of the criterion function $\phi_{c,\lambda}$ with respect to $\|\cdot\|_\infty$ that (2.16) holds for an arbitrary convex function g satisfying the side conditions $g(-c) = k_1$ and $g(c) = k_2$.

Conversely, suppose that f minimizes $\phi_c(g)$ over $\mathcal{G}(c, k_1, k_2)$. Let H be its second integral on $[-c, c]$, satisfying $H(-c) = Y(-c)$ and $H(c) = Y(c)$, and let let $F = H'$. If

$$g_{t,\epsilon}(u) = f(u) + \epsilon(u - t)_+ - \epsilon(c - t)(u + c)/(2c)$$

for $\epsilon > 0$ and $t \in (-c, c)$, then $g_{t,\epsilon}(-c) = k_1$, $g_{t,\epsilon}(c) = k_2$, and

$$H(t) - Y(t) = \lim_{\epsilon \downarrow 0} \frac{\phi_c(g_{t,\epsilon}) - \phi_c(f)}{\epsilon} \geq 0,$$

since f minimizes $\phi_c(g)$ over $\mathcal{G}(c, k_1, k_2)$. This yields (2.7). Again by the assumption that Y does not have one-sided parabolic tangents at $-c$ and c, we get from this that $F(-c) > X(-c)$ and $F(c) < X(c)$. This implies as before that f' has finite limits at $-c$ and c.

Next, taking

$$g_\epsilon(t) = f(t) + \epsilon f(t) - \epsilon k_1 - \epsilon\frac{(k_2-k_1)(t+c)}{2c},\ t \in [-c,c],$$

we again get $g_\epsilon(-c) = k_1$, $g_\epsilon(c) = k_2$, and by integration by parts and the finiteness of the limits of f' at $-c$ and c we obtain:

$$\int_{(-c,c)} \{H(t) - Y(t)\}\, df'(t) = \lim_{\epsilon\downarrow 0} \frac{\phi_c\left(g_\epsilon\right) - \phi_c(f)}{\epsilon} \geq 0.$$

and

$$-\int_{(-c,c)} \{H(t) - Y(t)\}\, df'(t) = \lim_{\epsilon\uparrow 0} \frac{\phi_c\left(f + g_\epsilon\right) - \phi_c(f)}{(-\epsilon)} \geq 0.$$

Hence

$$\int_{(-c,c)} \{H(t) - Y(t)\}\, df'(t) = 0,$$

yielding (2.8). Since (2.9) is also satisfied, we now also have proved the necessity of the conditions (2.7) to (2.9). □

An interesting property of the third derivative of the function H, satisfying the conditions of Lemma 2, is given in the following corollary.

Corollary 1.

(i) Suppose that the function H on $[-c,c]$ satisfies the conditions of Lemma 2. Then the third (left- or right-continuous) derivative $H^{(3)}$ of H is a bounded monotone increasing function that only grows on the "set of touch" S, defined by

$$S = \{t \in (-c,c) : H(t) = Y(t),\ H'(t) = X(t)\}. \tag{2.17}$$

The set S is closed and has Lebesgue measure zero.

(ii) With probability one, H is three times differentiable at zero.

Proof: ad (i) Since $H \geq Y$ and $\int_{(-c,c)} \{H(t) - Y(t)\}\, dH^{(3)}(t) = 0$, we must have:

$$\int_{\{t \in (-c,c): H(t) \neq Y(t)\}} dH^{(3)}(t) = 0.$$

Since a differentiable function has derivative zero at a relative minimum (see e.g. Dieudonné (1969), page 153, problem 3, part (a)), it follows that

$$\{t \in (-c,c) : H(t) = Y(t)\} = \{t \in (-c,c) : H(t) = Y(t),\, H'(t) = X(t)\}.$$

Since H has a bounded second derivative, there exists a constant $a > 0$ such that the function $\tilde{H}(t) = H(t) - at^2$ is concave on $[-c, c]$. Since the least concave majorant M of the function $\tilde{Y}(t) = Y(t) - at^2$ on $[-c, c]$ is the pointwise minimum of all concave functions lying above $\tilde{Y}$, we must have:

$$\tilde{H}(t) \geq M(t),\, t \in [-c, c],$$

and so $\tilde{H} \geq M \geq \tilde{Y}$. According to definition 1 and theorem 1 in Sinai (1992), the derivative of M decreases on a set with Lebesgue measure zero (a Cantor-type set). A point of touch of $\tilde{H}$ with $\tilde{Y}$ is necessarily a point of touch of M with $\tilde{Y}$. The set of locations of points of touch between H and Y is therefore a set with Lebesgue measure zero. The boundedness of $H^{(3)}$ follows again from the assumption that Y does not have one-sided parabolic tangents at $-c$ and c, implying $F(-c) > X(-c)$ and $F(c) < X(c)$, as in the proof of Lemma 2.

Finally, the set S is closed, since the function $H - Y$ is continuous on $[-c, c]$.

ad (ii) A fixed point will with probability zero belong to the Cantor-type set, described in (i), so in particular 0 will belong with probability zero to this set. This means that 0 is with probability zero the location of a point of touch of H and Y, and this in turn means that H'' has with probability zero a change of slope at 0. Since H'' is convex, it has left and right derivatives at zero, and since H'' has with probability zero a change of slope at 0, the right derivative cannot be bigger than the left derivative. □

The following lemma gives the structure of the function H of Lemma 2 on an "excursion interval" $[\tau_1, \tau_2]$ between two locations of points of touch τ_1 and τ_2 between H and Y. By "excursion interval" we mean that

$$H(\tau_1) = Y(\tau_1),\ H(\tau_2) = Y(\tau_2) \text{ and } H(t) > Y(t),\ t \in (\tau_1, \tau_2).$$

Note that such intervals exist by the construction in the proof of Corollary 1, where it was shown that the set of locations of points of touch between H and Y can be embedded (after a transformation) in the set of locations of points of touch of the concave majorant of drifting integrated Brownian motion.

LEMMA 3. *Suppose that the function H on $[-c, c]$ satisfies the conditions of Lemma 2 and that $[\tau_1, \tau_2]$ is an excursion interval for H w.r.t. Y, where $-c < \tau_1 < \tau_2 < c$. Let*

$$\overline{\tau} = \{\tau_1 + \tau_2\}/2,\ \overline{X} = \{X(\tau_1) + X(\tau_2)\}/2,\ \overline{Y} = \{Y(\tau_1) + Y(\tau_2)\}/2,$$

and

$$\Delta X = X(\tau_2) - X(\tau_1),\ \Delta Y = Y(\tau_2) - Y(\tau_1) \text{ and } \Delta\tau = \tau_2 - \tau_1.$$

Then the restriction of H to $[\tau_1, \tau_2]$ is given by

$$\begin{aligned} H(t) &= \frac{Y(\tau_2)(t-\tau_1) + Y(\tau_1)(\tau_2 - t)}{\Delta\tau} \\ &\quad - \tfrac{1}{2}\left\{\frac{\Delta X}{\Delta\tau} + \frac{4(\bar{X}\Delta\tau - \Delta Y)(t-\overline{\tau})}{(\Delta\tau)^3}\right\}(t-\tau_1)(\tau_2 - t),\ t \in [\tau_1, \tau_2]. \end{aligned} \tag{2.18}$$

The values of H, F at $\bar{\tau}$ are given by

$$H(\overline{\tau}) = \overline{Y} - \tfrac{1}{8}\Delta X \Delta\tau, \qquad F(\overline{\tau}) = H'(\overline{\tau}) = \frac{3\Delta Y - \overline{X}\Delta\tau}{2\Delta\tau}, \tag{2.19}$$

and the values of f and f' at $\bar\tau$ by

$$f(\overline{\tau}) = H''(\overline{\tau}) = \frac{\Delta X}{\Delta\tau}, \qquad f'(\overline{\tau}) = H^{(3)}(\overline{\tau}) = \frac{12(\overline{X}\Delta\tau - \Delta Y)}{(\Delta\tau)^3}. \tag{2.20}$$

Proof: Since the measure df' is zero on (τ_1, τ_2), the function f is linear on the interval $[\tau_1, \tau_2]$. This means that H behaves as a cubic polynomial on $[\tau_1, \tau_2]$ that is completely determined by the values of H and H' at the boundary points. By Corollary 1 we have:

$$H(\tau_1) = Y(\tau_1),\ H(\tau_2) = Y(\tau_2),\ H'(\tau_1) = X(\tau_1),\ \text{and } H'(\tau_2) = X(\tau_2). \tag{2.21}$$

But it is easily checked that the cubic polynomial, defined by (2.18), satisfies the boundary conditions (2.21). The relations (2.19) and (2.20) follow from this representation. □

In the following we are going to use properties of ordinary Brownian motion and integrated Brownian motion. Ordinary two-sided Brownian motion (without drift), originating from zero, will be denoted by W and its integral by V, where V is "pinned down" at zero: $V(0) = 0$. We then will use certain stationarity properties of the point process of points of touch between Y and the function H of Lemma 2, as $c \to \infty$. As a preparation to this, we reformulate the result of Lemma 3 in terms of the non-drifting processes V and W.

Corollary 2. *Suppose that the function H on $[-c, c]$ satisfies the conditions of Lemma 2 and that $[\tau_1, \tau_2]$ is an excursion interval for H w.r.t. Y, where $-c < \tau_1 < \tau_2 < c$. Let*

$$\overline{\tau} = \{\tau_1 + \tau_2\}/2,\ \overline{W} = \{W(\tau_1) + W(\tau_2)\}/2,\ \overline{V} = \{V(\tau_1) + V(\tau_2)\}/2,$$

and

$$\Delta W = W(\tau_2) - W(\tau_1),\ \Delta V = V(\tau_2) - V(\tau_1)\ \textit{and}\ \Delta\tau = \tau_2 - \tau_1.$$

Finally, let the function G be defined by

$$G(t) = H(t) - t^4,\ t \in [-c, c], \tag{2.22}$$

and let f_0 be defined by

$$f_0(t) = 12t^2,\ t \in I\!R. \tag{2.23}$$

Then the value of $f - f_0$ at $\bar{\tau}$ is given by

$$f(\bar{\tau}) - f_0(\bar{\tau}) = G''(\bar{\tau}) = \frac{\Delta W}{\Delta \tau} + (\Delta\tau)^2, \tag{2.24}$$

and $f' - f_0'$ at $\bar{\tau}$ by

$$f'(\bar{\tau}) - f_0'(\bar{\tau}) = G^{(3)}(\bar{\tau}) = \frac{12(\overline{W}\Delta\tau - \Delta V)}{(\Delta\tau)^3}. \tag{2.25}$$

The function $f - f_0$ has the following representation on $[\tau_1, \tau_2]$:

$$f(t) - f_0(t) = (\Delta\tau)^2 + \frac{\Delta W}{\Delta\tau} + \frac{12(t-\bar{\tau})\int_{\tau_1}^{\tau_2}(u-\bar{\tau})\,dW(u)}{(\Delta\tau)^3} - 12(t-\bar{\tau})^2. \tag{2.26}$$

Proof: This follows easily from Lemma 3. □

We will need the following two lemmas for the existence of a process, satisfying the conditions (i) to (iii) at the beginning of this section.

LEMMA 4. *Let, for each $c > 0$, H_c be the function, satisfying the conditions of Lemma 2, with $k_1 = k_2 = 12c^2$ and let t be a fixed point in $(-c, c)$. Furthermore, let $\tau_1 \le t$ be the location of the last point of touch between H_c and Y on $[-c, t]$ (note that, with probability one, $\tau_1 \ne t$) and let $\tau_2 > t$ be the location of the first point of touch between H_c and Y on $(t, c]$. Then, for every $\epsilon > 0$, there is an $M = M_\epsilon$ so that*

$$\limsup_{c\to\infty} P\left\{\tau_1 < t - M,\ \tau_2 > t + M\right\} \le \epsilon.$$

Proof: We first consider the special case $t = 0$. The cubic polynomial P_c such that $P_c(-c) = Y(-c)$, $P_c(c) = Y(c)$, and $P_c''(-c) = P_c''(c) = 12c^2$, is given by

$$P_c(t) = \frac{1}{2}(Y(-c) + Y(c)) - 6c^4 + \frac{1}{2}(Y(c) - Y(-c))\frac{t}{c} + 6c^2t^2$$

Hence,

$$P\{P_c(0) \geq 0\} = P\{Y(-c)+Y(c) \geq 12c^4\} = P\{V(-c)+V(c) \geq 10c^4\} \to 0 \text{ as } c \to \infty$$

since $V(\pm c) = \mathcal{O}_P(c^{3/2})$.

This means that the probability that the function H_c will at least have one point of touch with Y, apart from $-c$ and c, tends to 1, as $c \to \infty$, since we must have $H_c(0) \geq Y(0) = 0$ (note, as in the proof of Lemma 3, that Corollary 1 implies that $f_c = H_c''$ is linear on regions where H_c and Y do not touch, so H_c behaves as a cubic polynomial on such regions).

For similar reasons the probability that there will be both a point of touch in the interval $(-c, 0)$ and a point of touch in the interval $(0, c)$ will tend to one, as $c \to \infty$. So we may assume that $\tau_1 \in (-c, 0)$ and $\tau_2 \in (0, c)$. This implies by (2.19) and the property $H \geq Y$,

$$Y(\overline{\tau}) \leq H_c(\overline{\tau}) = \overline{Y} - \tfrac{1}{8}\Delta X \Delta\tau,$$

which can be rewritten as

$$V(\overline{\tau}) + \overline{\tau}^4 \leq \tfrac{1}{2}\left\{V(\tau_1) + V(\tau_2)\right\} + \tfrac{1}{2}\left\{\tau_1^4 + \tau_2^4\right\} - \tfrac{1}{8}\left\{W(\tau_2) - W(\tau_1) + 4\tau_2^3 - 4\tau_1^3\right\}\Delta\tau$$

$$= \tfrac{1}{2}\left\{V(\tau_1) + V(\tau_2)\right\} - \tfrac{1}{8}\left\{W(\tau_2) - W(\tau_1)\right\}\Delta\tau + \tfrac{1}{2}\tau_1\tau_2\left\{\tau_1^2 + \tau_2^2\right\}.$$

Hence

$$P\{\tau_1 < -M,\, \tau_2 > M\}$$

$$= P\{Y(\overline{\tau}) \leq H_c(\overline{\tau}),\, \tau_1 < -M,\, \tau_2 > M\}$$

$$= P\left\{V(\overline{\tau}) - \overline{V} + \tfrac{1}{8}\Delta W \Delta\tau \le \tfrac{1}{2}\tau_1\tau_2(\tau_1^2+\tau_2^2) - \overline{\tau}^4,\ \tau_1 < -M,\ \tau_2 > M\right\}$$

$$\le P\Big\{V(\overline{s}) - \tfrac{1}{2}(V(s_1)+V(s_2)) + \tfrac{1}{8}(W(s_2)-W(s_1))\Delta s \le \tfrac{1}{2}s_1 s_2(s_1^2+s_2^2) - \overline{s}^4 \quad \text{for some } s_1 < -M,\ s_2 > M\Big\}$$

$$= P\left\{V(\overline{s}) - \tfrac{1}{2}(V(s_1)+V(s_2)) + \tfrac{1}{8}(W(s_2)-W(s_1))\Delta s \le -\frac{1}{16}(s_2-s_1)^4 \quad \text{for some } s_1 < -M,\ s_2 > M\right\}, \tag{2.27}$$

where $\overline{V} = \{V(\tau_1)+V(\tau_2)\}/2$, $\overline{s} = \{s_1+s_2\}/2$, and $\Delta s = s_2 - s_1$. Now we rewrite the process appearing in the last display:

$$\begin{aligned} &V(\overline{s}) - \tfrac{1}{2}(V(s_1)+V(s_2)) + \tfrac{1}{8}(W(s_2)-W(s_1))\Delta s \\ &= -\tfrac{1}{2}\left\{V(s_2) - V(\overline{s}) - (s_2-\overline{s})W(\overline{s}) - \tfrac{1}{2}(W(s_2)-W(\overline{s}))(s_2-\overline{s})\right\} \\ &\qquad -\tfrac{1}{2}\left\{V(s_1) - V(\overline{s}) - (s_1-\overline{s})W(\overline{s}) - \tfrac{1}{2}(W(s_1)-W(\overline{s}))(s_1-\overline{s})\right\} \\ &\equiv -\tfrac{1}{2}\left\{\theta_2(\overline{s},s_2) + \theta_1(\overline{s},s_1)\right\} \end{aligned}$$

with θ_1, θ_2 defined by the last equality. Note that the process $\{Z(t),\mathcal{F}_t\}_{t\ge 0} \equiv \{V(t) - tW(t),\mathcal{F}_t\}_{t\ge 0}$, with $\mathcal{F}_t = \sigma\{W(s):\ 0 \le s \le t\}$, is a zero-mean martingale. Moreover, $Z(t) = -\int_0^t s\,dW(s)$ and hence $E\{Z(t)^2\} = t^3/3$. Similarly, $\{V(-t) + tW(-t),\mathcal{G}_t\}_{t\ge 0}$, with $\mathcal{G}_t = \sigma\{W(-s):\ 0 \le s \le t\}$, is a martingale. Hence, using a symmetry argument for $\theta_1(\overline{s},s_1)$ and $\theta_2(\overline{s},s_2)$, it is seen that the probability in (2.27) is bounded by

$$\begin{aligned} &4P\left\{V(t) - \tfrac{1}{2}tW(t) \ge t^4 \text{ for some } t > M\right\} \\ &\le 4P\left\{|V(t) - tW(t)| \ge \tfrac{1}{2}t^4 \text{ for some } t > M\right\} + 4P\left\{|\tfrac{1}{2}tW(t)| \ge \tfrac{1}{2}t^4 \text{ for some } t > M\right\} \\ &\le 4\sum_{j=[M]}^{\infty} \frac{(j+1)^3}{(j^4/2)^2} + 4\sum_{j=[M]}^{\infty}\frac{(j+1)}{(j^3)^2} \\ &\le C\sum_{j=[M]}^{\infty} j^{-5} \to 0 \qquad \text{as} \qquad M\to\infty, \end{aligned}$$

for some absolute constant C.

The statement for general t is proved along similar lines, conditioning on the value of the processes X and Y at the point t. □

LEMMA 5. *For each $c > 0$, let H_c be the function satisfying the conditions of Lemma 2, with $k_1 = k_2 = 12c^2$. Let f_c be the second derivative of H_c on $[-c, c]$. Then, for $t \in I\!R$ fixed, the collections $\{f_c(t) - f_0(t)\}_{c>|t|}$, $\{f_c^-(t) - f_0'(t)\}_{c>|t|}$, and $\{f_c^+(t) - f_0'(t)\}_{c>|t|}$ are tight; here f_c^+ and f_c^- denote the right and left derivatives of the convex function f_c.*

Proof: We prove the statement for the case $t = 0$, since the general statement for arbitrary t is proved in an entirely similar way, but involves more cumbersome notation. Let $\epsilon > 0$ and let $F_c = H_c'$. By Lemma 4, there exists for c large at least one point of touch, τ_2 say, in the interval $[-M, M]$, with probability at least $1 - \epsilon$, if $M < c$ is sufficiently large. Without loss of generality, suppose that $0 \le \tau_2 \le M$. By repeating the argument in Lemma 4 we can find another point of touch, τ_1 say, between $-3M$ and $-M$, perhaps at the cost of increasing M. Then by the mean value theorem it follows that for some $\xi_1 \in [\tau_1, \tau_2] \subset [-3M, M]$

$$f_c(\xi_1) = \frac{F_c(\tau_2) - F_c(\tau_1)}{\tau_2 - \tau_1} = \frac{X(\tau_2) - X(\tau_1)}{\tau_2 - \tau_1}$$

which is tight by virtue of Lemma 4, the construction of τ_1, τ_2, and by the definition of $X(t) = W(t) + 4t^3$.

Suppose that $\xi_1 < 0$. By repeating the above argument we can find another point of touch $\tau_3 \in (2M, 4M]$ and another point $\xi_2 \in [\tau_2, \tau_3] \subset [0, 4M]$ with

$$f_c(\xi_2) = \frac{F_c(\tau_3) - F_c(\tau_2)}{\tau_3 - \tau_2} = \frac{X(\tau_3) - X(\tau_2)}{\tau_3 - \tau_2}$$

which is again tight. Since f_c is convex it follows that

$$f_c(0) = f_c(\lambda\xi_1 + (1-\lambda)\xi_2) \le \lambda f_c(\xi_1) + (1-\lambda) f_c(\xi_2)$$

with $\lambda = \xi_2/(\xi_2 - \xi_1) \in [0,1]$. Since the right side is tight, this completes the proof of tightness of $\{f_c(0)\}$ assuming that $\xi_1 < 0$, since we can the use the argument of the first paragraph of the proof of Lemma 1 again for the lower bound of $f_c(0)$.

If $\xi_1 \ge 0$, we repeat the above argument to the left of zero again to find another $\xi_2 < 0$ with $f_c(\xi_2)$ tight again, and again conclude that $\{f_c(0)\}$ is tight.

Now suppose that we have produced points ξ_1 and ξ_2 with $-c < \xi_1 < -M < 0 < M < \xi_2 < c$ and $\{f_c(\xi_i)\}$ tight, $i = 1, 2$. Then, since all lines of slope $s \in [f_c^-(0), f_c^+(0)]$ lie below f_c, it follows that

$$f_c(\xi_2) \ge s\xi_2 + f_c(0) \ge sM + f_c(0)$$

for any $s \in [0, f_c^+(0) \vee 0]$. Thus it follows that

$$f_c^+(0) \le 0 \vee \frac{f_c(\xi_2) - f_c(0)}{M} \tag{2.28}$$

where the right side is tight. Similarly, using the point $\xi_1 < -M$, we find that

$$f_c^-(0) \ge 0 \wedge -\frac{f_c(\xi_1) - f_c(0)}{M} \tag{2.29}$$

where the right side is tight. Combining (2.28) and (2.29) yields the conclusion for $\{f_c^-(0)\}$ and $\{f_c^+(0)\}$. □

We now define the collection of convex functions f_n on $[-n, n]$ as the second derivatives of the functions H_n, satisfying the conditions of Lemma 2, with $k_1 = k_2 = 12n^2$, and extend these functions to $\mathbb{R}$ by linearly extending them from $-n$ and n, respectively. On a set with probability one the possibility of such an extension exists, since we may assume that Y has no parabolic tangents at $-n$ and n, and hence that f_n has finite derivatives at $-n$ and n. The functions H_n and $F_n = H_n'$

are also continuously extended to functions on $I\!R$, by taking F_n and H_n as the first and second integral of f_n, respectively, uniquely determined by their values at the points $-n$ and n, where we start the extension.

Moreover, we define, for each $M > 0$, the seminorms

$$\|H\|_M = \sup_{t \in [-M,M]} \{|H(t)| + |H'(t)| + |H''(t)|\} \tag{2.30}$$

on the set of twice continuously differentiable functions $H : I\!R \to I\!R$. We now have the following result.

COROLLARY 3. *Let $X(t) = W(t) + 4t^3$ where $W(t)$ is standard two-sided Brownian motion starting from 0, and let Y be the integral of X, satisfying $Y(0) = 0$. Then almost surely there exists a continuous stochastic process H (the "invelope"), satisfying the conditions (i) to (iii) at the beginning of this section.*

Proof. We show that the sequence (H_n), where H_n is defined as in Lemma 5, with c replaced by n, and continuously extended to functions on $I\!R$ as second integrals of the linearly extended functions f_n (as indicated above), has a convergent subsequence in the topology induced by the semi-norms (2.30).

Fix $m > 0$ in $\mathbb{N}$. Let τ_n^+ be the location of the first point of touch $\geq m$ between H_n and Y and let τ_n^- be the location of the last point of touch $\leq -m$ between H_n and Y. Since the set of the locations of points of touch is closed according to Corollary 1 and with probability one not empty by Lemma 4, such points exist, for sufficiently large $n > m$. Moreover, by Lemma 4, the sequences (τ_n^-) and (τ_n^+) are almost surely bounded, so they have convergent subsequences $(\tau_{n_k}^-)$ and $(\tau_{n_k}^+)$ such that, almost surely,

$$\lim_{k\to\infty} \tau_{n_k}^- = \tau^- \text{ and } \lim_{k\to\infty} \tau_{n_k}^+ = \tau^+,$$

say, where $\tau^-,\tau^+\in I\!R$. Since $H_n(\tau_n^-)=Y(\tau_n^-)$, and $H_n(\tau_n^+)=Y(\tau_n^+)$, and Y is continuous, this means:

$$H_{n_k}\left(\tau_{n_k}^-\right)\to Y(\tau^-) \text{ and } H_{n_k}\left(\tau_{n_k}^+\right)\to Y(\tau^+),$$

as $k\to\infty$. Similarly, since also X is continuous,

$$H'_{n_k}\left(\tau_{n_k}^-\right)\to X(\tau^-) \text{ and } H'_{n_k}\left(\tau_{n_k}^+\right)\to X(\tau^+),.$$

as $k\to\infty$.

Suppose $M>0$ satisfies

$$-M<\tau^-<\tau^+<M.$$

and let $f_n=H_n''$. By Lemma 5, the collections $\{f_n(t)-f_0(t)\}_{n>M}$, $\{f_n^+(t)-f_0'(t)\}_{n>M}$ and $\{f_n^-(t)-f_0'(t)\}_{n>M}$ are tight, for $t=0$ and $t=\pm M$, so we may assume that these sequences are bounded. This means, by the monotonicity of f_n^+ and f_n^-, that the functions f_n have uniformly bounded derivatives on $[-M,M]$. So, by the Arzelà-Ascoli theorem, the sequence of functions (f_{n_k}), restricted to $[-M,M]$, has a subsequence (f_{n_ℓ}), converging in the supremum metric on continuous functions on $[-M,M]$ to a bounded convex function $f:[-M,M]\to I\!R$. Since the functions $(f_{n_\ell}|[-M,M])$ are uniformly bounded, we can now also apply the Arzelà-Ascoli theorem to the uniformly bounded sequence $(F_{n_\ell}|[-m,m])$, where $F_n=H_n'$, to conclude that this sequence has a convergent sequence in the supremum metric of continuous functions on $[-m,m]$. Finally, repeating the argument for H_n itself, we find that there is a further subsequence (n_j) such that $(H_{n_j}|[-m,m])$ converges in the supremum metric of continuous functions on $[-m,m]$.

Thus, starting with the sequence (H_n) we can find a subsequence (H_{n_j}) so that $(H_{n_j}|[-m,m])$ converges in the topology induced by the metric $\|H\|_m$ to a limit function $H^{(m)}$ with convex second derivative $f^{(m)}$ on $[-m, m]$. By a diagonal argument we now get that the sequence (H_n) has a subsequence (H_{n_k}) converging in the topology induced by the semi-norms $\|H\|_m$, $m = 1, 2, \ldots$ to a function H with convex second derivative f. It is clear that this limit H satisfies the conditions (i) to (iii) of Theorem 1. □

We still have to show that if two functions G and H both satisfy conditions (i) to (iii) of Theorem 1, they must be equal with probability 1. To this end we first prove that if G and H have two different common points of touch $a < b$ with Y, they must be equal on the interval $[a, b]$.

Lemma 6. *Suppose that G and H both satisfy conditions (i) to (iii) of Theorem 1. If G and H have two common points of touch with Y at a and b, where $a < b$, then $G \equiv H$ on $[a, b]$.*

Proof. Let $g = G''$ and $h = H''$, and let, for a convex function f on $[a, b]$, $\phi_{a,b}(f)$ be defined by

$$\phi_{a,b}(f) = \tfrac{1}{2}\int_a^b f(t)^2\,dt - \int_a^b f(t)\,dX(t). \tag{2.31}$$

Then we get:

$$\phi_{a,b}(g) - \phi_{a,b}(h) = \tfrac{1}{2}\int_a^b \{h(t) - g(t)\}^2\,dt + \int_a^b \{H(t) - Y(t)\}\,dg'(t). \tag{2.32}$$

This is seen as follows. Using (2.11) it follows that

$$\begin{aligned}\phi_{a,b}(g) - \phi_{a,b}(h) &= \tfrac{1}{2}\int_a^b \{g(t) - h(t)\}^2\,dt + \int_a^b \{g(t) - h(t)\}\,h(t)\,dt \\ &\qquad - \int_a^b \{g(t) - h(t)\}\,dX(t) \\ &= \tfrac{1}{2}\int_a^b \{g(t) - h(t)\}^2\,dt - \int_a^b \{g'(t) - h'(t)\}\,\{H'(t) - X(t)\}\,dt\end{aligned}$$

$$= \tfrac{1}{2}\int_a^b \{g(t)-h(t)\}^2\,dt + \int_a^b \{H(t)-Y(t)\}\,d\{g'-h'\}(t)$$
$$= \tfrac{1}{2}\int_a^b \{g(t)-h(t)\}^2\,dt + \int_a^b \{H(t)-Y(t)\}\,dg'(t),$$

using $H(a)=Y(a)$, $H'(a)=X(a)$ and similar equalities at the point b (a and b are points of touch for H and Y and H' must also be equal to X at these points, because $H\ge Y$ and 2.3). Similarly, we get

$$\phi_{a,b}(h)-\phi_{a,b}(g) = \tfrac{1}{2}\int_a^b \{g(t)-h(t)\}^2\,dt + \int_a^b \{G(t)-Y(t)\}\,dh'(t). \tag{2.33}$$

Since the right-hand sides of (2.32) and (2.33) are nonnegative, we must have $\phi_{a,b}(g)=\phi_{a,b}(h)$ and hence $g\equiv h$ on $[a,b]$. Moreover, since a and b are points of touch of G and Y and of H and Y, we have:

$$G(a)=H(a)=Y(a),\; G'(a)=H'(a)=X(a),$$

and

$$G(b)=H(b)=Y(b),\; G'(b)=H'(b)=X(b).$$

Hence also $G\equiv H$ on $[a,b]$. □

We will also need the following lemma.

Lemma 7. *Suppose H is a function, satisfying conditions (i) to (iii) of Theorem 1, with second derivative h. Let $t\in I\!R$ and $\epsilon>0$, and let τ_- be the location of the last point of touch $\le t$ of H and Y and let τ_+ be the location of the first point of touch $>t$ of H and Y. Furthermore, let $f_0: I\!R\mapsto I\!R$ be defined by $f_0(t)=12t^2$. Then we have the following properties:*

(i) There exists an $M=M(\epsilon)>0$, independent of t, such that

$$P\{(t-\tau_-)\vee(\tau_+-t)>M\}<\epsilon. \tag{2.34}$$

(ii) There exists an $M = M(\epsilon) > 0$, independent of t, such that

$$P\{|h(t) - f_0(t)| > M\} < \epsilon, \tag{2.35}$$

$$P\left\{\left|h^+(t) - f_0'(t)\right| > M\right\} < \epsilon, \tag{2.36}$$

and

$$P\left\{\left|h^-(t) - f_0'(t)\right| > M\right\} < \epsilon, \tag{2.37}$$

where h^+ and h^- denote the right and left derivatives of h, respectively.

Proof: The proof follows the same pattern as the proof of Lemma's 4 and 5, and uses the stationarity of the increments of W and the integrated Brownian motion process (without drift) V. For example, if $[\tau_-, \tau_+]$ is an "excursion interval for H", we have the representation

$$h(t) - f_0(t) = (\Delta\tau)^2 + \frac{\Delta W}{\Delta\tau} + \frac{12(t-\overline{\tau})\int_{\tau_-}^{\tau_+}(u-\overline{\tau})\,dW(u)}{(\Delta\tau)^3} - 12(t-\overline{\tau})^2. \tag{2.38}$$

on $[\tau_-, \tau_+]$, just as (2.26) in Corollary 2, where $\bar{\tau} = (\tau_- + \tau_+)/2$, $\Delta W = W(\tau_+) - W(\tau_-)$, and $\Delta\tau = \tau_+ - \tau_-$.

Part (i) follows from the inequality

$$\begin{aligned}
&P\{\tau_- < t - M,\, \tau_+ > t + M\} \\
&\le 4P\left\{V(u) - V(t) - \tfrac{1}{2}(u-t)\{W(u) - W(t)\} \ge \tfrac{1}{2}(u-t)^4,\ \text{for some } u > t + M\right\} \\
&\le C\sum_{j=[M]}^{\infty} j^{-5},
\end{aligned}$$

for some absolute constant $C > 0$, similarly to (2.27). The stationarity of the increments of W and V implies that the upper bound is independent of t.

Part (ii) is proved along the lines of the proof of Lemma 5. □

We are now ready to prove Theorem 1.

Proof of Theorem 1: Existence of a function H satisfying (i) to (iii) of Theorem 1 follows from Corollary 3. So the only remaining task is to prove uniqueness of this function H. So suppose that G and H, with second derivatives g and h, satisfy (i) to (iii) of Theorem 1, and that $G \neq H$.

Lemma 6 implies that, if $G \neq H$ on an interval $[a,b]$, there cannot be points $a' < a$ and $b' > b$ such that G and H have common points of touch with Y at a' and b', since in that case $G \equiv H$ on $[a',b']$ and hence also $G \equiv H$ on $[a,b]$, since $[a,b] \subset [a',b']$. This means that, if $G \neq H$ on an interval $[a,b]$, either *all* points of touch between G and Y at points $b' > b$ are different from *all* points of touch between H and Y at locations to the right of b or *all* points of touch between G and Y at points $a' < a$ are different from *all* points of touch between H and Y at locations to the left of a (or both).

First suppose that $G \neq H$ on an interval $[a,b]$ and that *all* points of touch between G and Y at points $b' > b$ are different from *all* points of touch between H and Y at locations to the right of b and *all* points of touch between G and Y at points $a' < a$ are different from *all* points of touch between H and Y at locations to the left of a (we will look at the "one-sided situation" at the end of the proof).

Let, for each n, a_n^G be the location of the first point of touch between G and Y to the left of $-n$, and b_n^G be the location of the first point of touch between G and Y to the right of n. Furthermore, let, for each n, a_n^H be the location of the first point of touch between H and Y to the left of a_n^G, and b_n^H be the location of the first point of touch between H and Y to the right of b_n^G. By assumption, $a_n^G \neq a_n^H$ and $b_n^G \neq b_n^H$ for sufficiently large n. Note that such "first points" exist, since, by Corollary 1 and Lemma 7 (i), the set of locations of points of touch is closed and non-empty with probability one.

Finally, let, for a convex function f on $[a,b]$, $\phi_{a,b}(f)$ be defined as in (2.31):

$$\phi_{a,b}(f) = \tfrac{1}{2}\int_a^b f(t)^2\,dt - \int_a^b f(t)\,dX(t). \tag{2.39}$$

Then we have:

$$\begin{aligned} &\phi_{a_n^H,b_n^H}(g) - \phi_{a_n^H,b_n^H}(h) \\ &\quad = \tfrac{1}{2}\int_{a_n^H}^{b_n^H} \{h(t)-g(t)\}^2\,dt + \int_{a_n^H}^{b_n^H} \{H(t)-Y(t)\}\,dg'(t)\,, \end{aligned} \tag{2.40}$$

and similarly:

$$\begin{aligned} &\phi_{a_n^G,b_n^G}(h) - \phi_{a_n^G,b_n^G}(g) \\ &\quad = \tfrac{1}{2}\int_{a_n^G}^{b_n^G} \{g(t)-h(t)\}^2\,dt + \int_{a_n^G}^{b_n^G} \{G(t)-Y(t)\}\,dh'(t), \end{aligned} \tag{2.41}$$

see (2.32) and (2.33). Addition of (2.40) and (2.41) yields:

$$\begin{aligned} &\phi_{a_n^H,b_n^H}(g) - \phi_{a_n^H,b_n^H}(h) + \phi_{a_n^G,b_n^G}(h) - \phi_{a_n^G,b_n^G}(g) \\ &= \tfrac{1}{2}\int_{a_n^G}^{b_n^G} \{g(t)-h(t)\}^2\,dt + \tfrac{1}{2}\int_{a_n^H}^{b_n^H} \{h(t)-g(t)\}^2\,dt \\ &\qquad + \int_{a_n^H}^{b_n^H} \{H(t)-Y(t)\}\,dg'(t) + \int_{a_n^G}^{b_n^G} \{G(t)-Y(t)\}\,dh'(t) \\ &= \tfrac{1}{2}\int_{J_n\cup K_n} \{g^2(t)-h^2(t)\}\,dt - \int_{J_n\cup K_n} \{g(t)-h(t)\}\,dX(t) \\ &= \tfrac{1}{2}\int_{J_n\cup K_n} \{g(t)-h(t)\}\{g(t)-f_0(t)\}\,dt + \tfrac{1}{2}\int_{J_n\cup K_n} \{g(t)-h(t)\}\{h(t)-f_0(t)\}\,dt \\ &\qquad\qquad - \int_{J_n\cup K_n} \{g(t)-h(t)\}\,dW(t) \\ &\geq \int_{-n}^{n} \{g(t)-h(t)\}^2\,dt \end{aligned} \tag{2.42}$$

for all large n where $J_n = \left[a_n^H, a_n^G\right]$, $K_n = \left[b_n^G, b_n^H\right]$, and $f_0(t) = 12t^2$.

Now first suppose:

$$\lim_{n\to\infty}\int_{-n}^{n} \{g(t)-h(t)\}^2\,dt < \infty.$$

Then

$$\lim_{t\to-\infty}\{g(t)-h(t)\}=0 \text{ and } \lim_{t\to\infty}\{g(t)-h(t)\}=0.$$

This implies that

$$\liminf_{n\to\infty}\int_{J_n\cup K_n}\{g(t)-h(t)\}\,dW(t)=0,$$

almost surely, since, by Lemma 7(i), the lengths of J_n and K_n are $O_p(1)$, uniformly in n, and since g, h are continuous, implying that for each $\epsilon>0$:

$$P\left\{\liminf_{n\to\infty}\int_{J_n\cup K_n}\{g(t)-h(t)\}\,dW(t)>\epsilon,\ \lim_{|t|\to\infty}\{g(t)-h(t)\}=0\right\}$$
$$=\lim_{n\to\infty}P\left\{\inf_{i\ge n}\int_{J_i\cup K_i}\{g(t)-h(t)\}\,dW(t)>\epsilon,\ \lim_{|t|\to\infty}\{g(t)-h(t)\}=0\right\}$$
$$\le\lim_{n\to\infty}P\left\{\int_{J_n\cup K_n}\{g(t)-h(t)\}\,dW(t)>\epsilon,\ \lim_{|t|\to\infty}\{g(t)-h(t)\}=0\right\}=0.$$

This follows from the fact that, for example on the interval K_n, at least one of the two functions g and h has to be linear, implying that g and h can have at most two crossings on this interval, and possibly a region where they coincide, by the convexity of g and h. But this means that $g-h$ is a function of uniformly bounded variation on the interval K_n, with a supremum distance on this interval that tends to zero, as $n\to\infty$. A similar statement holds for the interval J_n. Since the length of the interval J_n, resp. K_n, is $\mathcal{O}_p(1)$, it follows that the limit in the last line in the above display has to be zero.

We similarly get, using the Cauchy-Schwarz inequality and Lemma 7(ii), that for each $\epsilon>0$ there exists a $\delta>0$ such that

$$P\left\{\liminf_{n\to\infty}\left\{\int_{J_n\cup K_n}\{g(t)-h(t)\}\{g(t)-f_0(t)\}\,dt\right\}^2>\epsilon,\ \lim_{|t|\to\infty}\{g(t)-h(t)\}=0\right\}$$
$$\le P\left\{\liminf_{n\to\infty}\int_{J_n\cup K_n}\{g(t)-h(t)\}^2\,dt\int_{J_n\cup K_n}\{g(t)-f_0(t)\}^2\,dt>\epsilon,\ \lim_{|t|\to\infty}\{g(t)-h(t)\}=0\right\}$$
$$=\lim_{n\to\infty}P\left\{\inf_{i\ge n}\int_{J_i\cup K_i}\{g(t)-h(t)\}^2\,dt\int_{J_i\cup K_i}\{g(t)-f_0(t)\}^2\,dt>\epsilon,\ \lim_{|t|\to\infty}\{g(t)-h(t)\}=0\right\}$$

$$\le \limsup_{n\to\infty} P\left\{\delta\int_{J_n\cup K_n}\{g(t)-f_0(t)\}^2\,dt>\epsilon\right\}<\epsilon\,.$$

A similar relation holds for

$$\left\{\int_{J_n\cup K_n}\{g(t)-h(t)\}\{h(t)-f_0(t)\}\,dt\right\}^2\,.$$

Thus

$$P\left\{\liminf_{n\to\infty}\int_{J_n\cup K_n}\{g(t)-h(t)\}\{g(t)-f_0(t)\}\,dt>0,\ \lim_{|t|\to\infty}\{g(t)-h(t)\}=0\right\}=0\,,$$

and similarly

$$P\left\{\liminf_{n\to\infty}\int_{J_n\cup K_n}\{g(t)-h(t)\}\{h(t)-f_0(t)\}\,dt>0,\ \lim_{|t|\to\infty}\{g(t)-h(t)\}=0\right\}=0.$$

But then (2.42) cannot hold for all large n, since

$$\int_{-n}^{n}\{g(t)-h(t)\}^2\ dt$$

tends to a strictly positive limit, as $n\to\infty$, if $h\ne g$.

Next, if

$$\lim_{n\to\infty}\int_{-n}^{n}\{g(t)-h(t)\}^2\ dt=\infty\,,$$

we also get a contradiction, using Lemma 7(ii), since

$$\begin{aligned}&\liminf_{n\to\infty}\int_{J_n\cup K_n}\{g(t)-h(t)\}\ dW(t)\\ &\qquad=\liminf_{n\to\infty}\int_{J_n\cup K_n}\{g(t)-f_0(t)-(h(t)-f_0(t))\}\ dW(t)<\infty,\end{aligned}$$

almost surely, and also, by the Cauchy-Schwarz inequality,

$$\begin{aligned}&\liminf_{n\to\infty}\left\{\int_{J_n\cup K_n}\{g(t)-h(t)\}\{g(t)-f_0(t)\}\,dt\right\}^2\\ &\le\liminf_{n\to\infty}\int_{J_n\cup K_n}\{g(t)-h(t)\}^2\,dt\int_{J_n\cup K_n}\{g(t)-f_0(t)\}^2\,dt<\infty,\end{aligned}$$

almost surely where the finiteness holds by Lemma 7(i), with a similar relation for

$$\left\{\int_{J_n\cup K_n}\{g(t)-h(t)\}\{h(t)-f_0(t)\}\,dt\right\}^2\,.$$

So again (2.42) cannot hold for all large n, if $h \neq g$.

Finally, if, for example, there would be infinitely many common points of touch a_n for a sequence (a_n) such that $a_n \to -\infty$, we consider $\phi_{a,b_n^G}(g) - \phi_{a,b_n^G}(h)$ and $\phi_{a,b_n^H}(g) - \phi_{a,b_n^H}(h)$, where $a = a_H = a_G$ is such a common point of touch (to the left of such a point the functions have to be equal!), and then we get a contradiction in the same way, if we assume $G \neq H$. □

Corollary 4. *Let f_c minimize ϕ_c defined in (2.4) over the set $\mathcal{G}(c, k_1(c), k_2(c))$ defined in (2.6), where*

$$|k_1(c) - 12c^2| \vee |k_2(c) - 12c^2| \leq M, \text{ for some fixed } M > 0,$$

and let f_c be linearly extended to a function on $I\!R$ on the intervals $(-\infty, -c]$ and $[c, \infty)$. Then f_c converges almost surely to the second derivative of the invelope H of Y, in the topology of uniform convergence on compacta. In particular:

$$\lim_{c \to \infty} f_c(0) = H''(0),$$

almost surely.

Proof: The proof of Corollary 3 showed that, taking $c_n = n$, there exists a subsequence (n_k) such that the functions H_{n_k}, defined as in Lemma 5, and continuously extended to functions on $I\!R$ as second integrals of the linearly extended functions f_{n_k}, converge to an invelope H of Y, in the topology induced by the semi-norms (2.30). It is also clear from the proof that if we take the boundary conditions

$$f(n) = k_1(-n),\ f(n) = k_2(n),$$

where $|k_1(n) - 12n^2| \vee |k_2(n) - 12n^2| \leq M$, instead of the boundary condition $f(-n) = f(n) = 12n^2$, we also get that there exists a subsequence (n_k) such that the functions H_{n_k}, continuously extended to a function on $I\!R$ as second integrals of the linearly extended functions f_{n_k}, converge to an invelope H of Y, in the topology induced by the semi-norms (2.30).

But since the invelope H is almost surely uniquely defined, and since the argument can be repeated for any subsequence, we get that the original sequence (H_n), continuously extended to functions on $\mathbb{R}$ as second integrals of the linearly extended functions f_n, also converges to the invelope H, in the topology induced by the semi-norms (2.30). Since the choice of $c_n = n$ is also irrelevant for the argument, we get that for any sequence (c_n) such that $c_n \to \infty$, the continuously extended functions H_{c_n} converge to H, again in the topology induced by the semi-norms (2.30).

This means that the continuously extended function H_c converges to H, as $c \to \infty$, in the topology induced by the semi-norms (2.30). By the definition of the semi-norms (2.30), this means that $f_c = H_c''$ converges to $f = H''$ in the topology of uniform convergence on compacta. □

Remark. Note that Corollary 4 shows that indeed the influence of the boundary conditions at $-c$ and c on the value of the function f_c in a fixed interval dies out, as $c \to \infty$, at least if we keep $f(-c) - f_0(-c)$ and $f(c) - f_0(c)$ bounded. But we got this result by using the unicity of the invelope H and not by directly comparing two solutions f_c and $\tilde{f}_c$ satisfying different boundary conditions at $-c$ and c, respectively. As noted in the introduction, comparing these solutions directly is difficult, since we cannot assume that the functions have changes of slope at the same points.

3. The iterative cubic spline algorithm The characterization of the solution of the minimization problem on a finite interval $[-c, c]$, given in Lemma 2, inspires an *iterative cubic spline algorithm* for finding the solution to the minimization problem of minimizing $\phi_c(g)$ over the set $\mathcal{G}(c, k_1, k_2)$ (for the notation, see (2.4) and (2.6)). For a full description of this algorithm in the finite sample problem, including a convergence proof for a general class of algorithms to which the iterative cubic spline algorithm belongs, we refer to GROENEBOOM, JONGBLOED, AND WELLNER (2001B).

The key idea behind the iterative cubic spline algorithm is the following. The minimizer of ϕ_c over the class of piecewise linear functions ℓ on $[-c, c]$ with set of knots $S = \{-c, t_1, t_2, \ldots, t_m, c\}$ satisfying $\ell(-c) = k_1$ and $\ell(c) = k_2$, is given by the second derivative of the cubic spline P that satisfies

$$P(t) = Y(t) \text{ for } t \in S,\ P''(-c) = k_1,\ P''(c) = k_2. \tag{3.1}$$

This can e.g. be seen by the arguments used in the proof of Lemma 2, or by direct differentiation.

Note that P'' satisfies a relation of the following type:

$$\begin{aligned} &\frac{t_j - t_{j-1}}{6} P''(t_{j-1}) + \frac{t_{j+1} - t_{j-1}}{3} P''(t_j) + \frac{t_{j+1} - t_j}{6} P''(t_{j+1}) \\ &= \frac{Y(t_{j+1}) - Y(t_j)}{t_{j+1} - t_j} - \frac{Y(t_j) - Y(t_{j-1})}{t_j - t_{j-1}}, \end{aligned} \tag{3.2}$$

for successive points t_{j-1}, t_j and t_{j+1}, see, e.g., (3.3.7) on p. 115 of PRESS ET AL. (1992).

The iterative cubic spline algorithm consists of two basic steps. The starting point at each iteration is a set of knots S together with a piecewise linear convex function f having set of knots S, that minimizes ϕ_c over the set of piecewise linear functions having the same set of knots. This means that the second integral H of f equals Y at points in S. If $H(t) \geq Y(t)$ for all $t \in [-c, c]$, the characterizaton of Lemma 2 shows that f is the solution of the minimization problem. If not, determine

$$t^* = \text{argmin}_{t \in [-c,c]} H(t) - Y(t)$$

and add this point to the present set of knots: $S := S \cup \{t^*\}$.

In the second step, the aim is to get a set of knots $S^* \subset S$ together with a *convex* piecewise linear function f having this set of knots such that f minimizes ϕ_c over the subset of $\mathcal{G}(c, k_1, k_2)$ consisting of functions with set of knots S^*. This is done by repeated computation of cubic splines as follows: The first cubic spline is the one defined by (3.1) with the extended set of knots S (including t^*). If P'' is convex, this iteration step is completed since P'' minimizes ϕ_c over the class $\mathcal{G}(c, k_1, k_2)$ consisting of functions with set of knots S. If P'' is not convex, we determine the maximal value of $\lambda \in (0, 1)$ for which $f + \lambda(P'' - f)$ is convex. Since $\lambda < 1$, this means that some knot in S actually vanishes. Removing this particular knot from S, we can again compute a cubic spline from (3.1) and check whether P'' is convex etc. Repeating this procedure, we get after finitely many (usually one or two) steps a set of knots S^* with corresponding P satisfying (3.1) such that P'' is convex. Then we turn to the first step of the next iteration again.

In Groeneboom, Jongbloed, and Wellner (2001b) it is shown that the iteration steps are well defined. Moreover, it is shown that the sequence of iterates f_n generated by the algorithm converges to the solution of the minimization problem.

The iterative cubic spline algorithm is directly motivated by the characterization of the solution of the minimization problem on a finite interval $[-c, c]$, given in Lemma 2. In that sense the algorithm is comparable to the convex minorant algorithm in the problem of estimation of a monotone function, which is also directly motivated by a geometric characterization of the solution of a minimization problem. The *hinge algorithm* introduced in Meyer (1997), can also be used to solve the minimization problem. The advantage of the iterative cubic spline algorithm compared to the hinge algorithm is that, in the computation of the splines, only a tridiagonal matrix has to be inverted (which can be seen from (3.2)), whereas the solution of the least squares problems in the hinge algorithm involves inversion of matrices that need not be tridiagonal.

A C program, implementing the iterative cubic spline algorithm was developed, and below we show some pictures of the "invelope" and its derivatives for solutions on the intervals $[-1,1]$ ($c=1$) and $[-4,4]$ ($c=4$), respectively. An approximation to Brownian motion on $[0,1]$ was generated with the Haar functions construction, see e.g., Rogers and Williams (1994), section 1.6. In the notation used there, we used the orthonormal functions

$$g_{k,n},\ k \leq 2^n,\ k \text{ odd},$$

up to $n=12$. The approximation to Brownian motion on $[-4,4]$ was generated by taking independent copies on the intervals $[i-1,i]$, $i=-3,\ldots,4$ and pasting these together at the borders of the intervals. Furthermore we took a grid of 8001 equidistant points on $[-4,4]$ and computed (an approximation to) the Brownian motion on these points.

In Figures 1 to 4 we compare the solution on $[-1,1]$ and $[-4,4]$, respectively, under the boundary conditions $f(-c)=f(c)=12c^2$. The functions with index 1 correspond to the solution for $c=1$ and the functions with index 2 to the solution for $c=4$. Figure 5 shows a comparison of the invelopes of two solutions for $c=1$, under the boundary conditions $f(-1)=f(1)=12$ and $f(-1)=f(1)=6$, respectively. Again the function with index 1 correspond to the solution for $c=1$ and $f(-1)=f(1)=12$ and the function with index 2 to the other solution.

The iterative cubic spline algorithm required (on a Macintosh powerbook 3400C) 11 iterations and less than 1 second for the solution for $c=1$ and $f(-1)=f(1)=12$, and 5 iterations and less than 1 second for the solution for $c=1$ and $f(-1)=f(1)=6$. The solution for $c=4$ took 45 iterations and 3 seconds.

This performance is pretty good in comparison with other algorithms we have tried (like the interior point method with logarithmic barrier function; see e.g. WRIGHT (1997)), in particular since for $c = 4$ a solution on a grid of 8001 points is needed. Figures 1 to 4 below illustrate the following facts.

1. The locations of the points of jump of the derivative of the solution change, as c increases. Note that the set of locations of points of jump of the derivative of the solution of the convex regression problem is the same as the set of locations of points of touch between the "invelope" and Y in the characterization of the solution in Lemma 2. For $c = 1$ we got the set of points $\{-0.931, -0.544, -0.116, 0.768\}$ and for $c = 4$ the set $\{-0.889, -0.886, -0.115, 0.616, 0.765\}$.
2. Figure 4 shows there is no evidence whatsoever that the changes of slope are bigger than a fixed constant (as claimed in WANG (1994)).
3. Figure 4 also shows that the derivative f_2', corresponding to the solution for $c = 4$ behaves better (in the sense that the absolute value of its difference with $f_0'(t)$ is smaller) at the boundary point -1 of the interval $[-1, 1]$ than the derivative $f_1'(-1)$ of the solution for $c = 1$. Phenomena like this are to be expected, since the solution on the interval $[-4, 4]$ poses more restrictions on the behavior of the solution on the smaller interval $[-1, 1]$. In fact, the tightness argument for $f_c^{-}(t) - f_0'(t)$ and $f_c^{+}(t) - f_0'(t)$, as $c \to \infty$ of Lemma 4 is partly illustrated here, at the point $t = -c$.

Further experiments showed that the solution on $[-1, 1]$ hardly changes if we increase c from 4 to, say, 5 or 6, in accordance with Corollary 4. Figure 5 shows that the locations of points of jump of the derivative of the solution of the convex regression problem (= the set of locations of points of touch between the "invelope" and

Y) change if we change the boundary condition on the value of f at -1 and 1. In this case we get the set of points $\{-0.931, -0.544, -0.116, 0.768\}$ for the boundary conditions $f(-1) = f(1) = 12$ (see above) and the set $\{-0.540, 0.179, 0.134\}$ for the boundary conditions $f(-1) = f(1) = 6$.

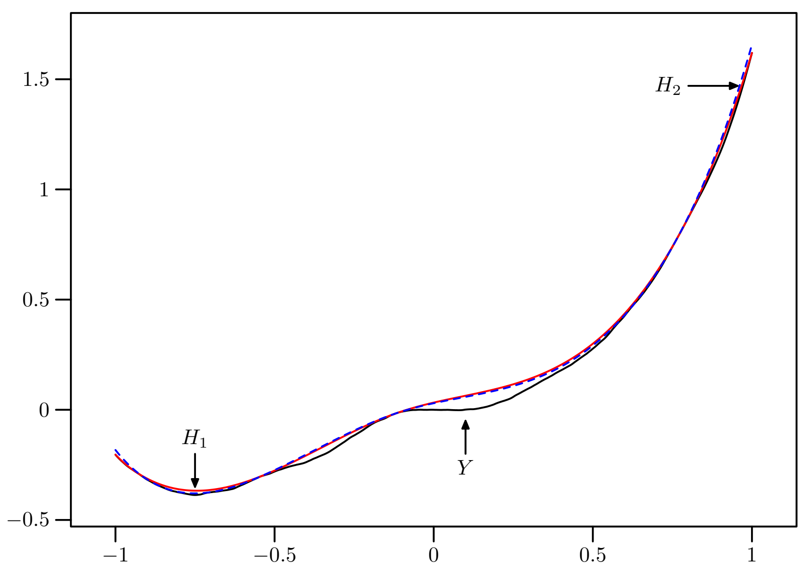

FIG. 1. *Solid: Y and H_1, dashed: H_2. Boundary conditions: $f_1(\pm 1) = 12$, $f_2(\pm 4) = 192$.*

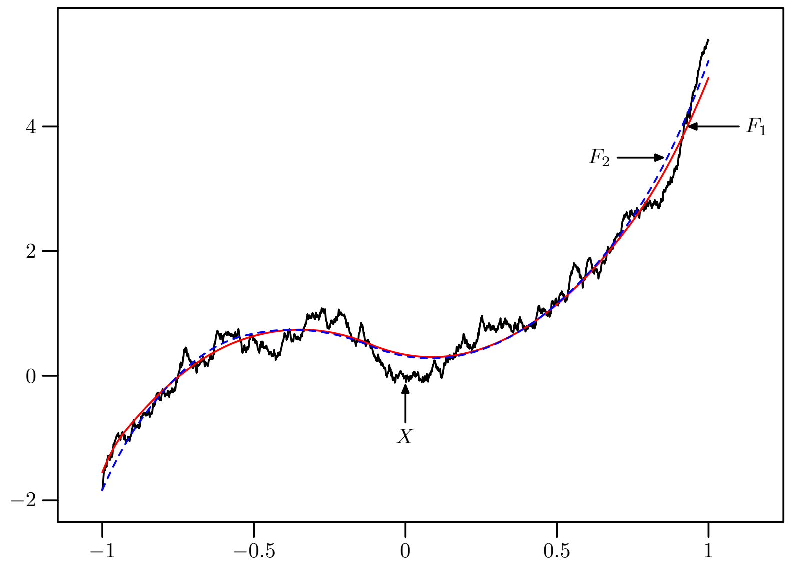

Fig. 2. *Solid: X and F_1, dashed: F_2.*

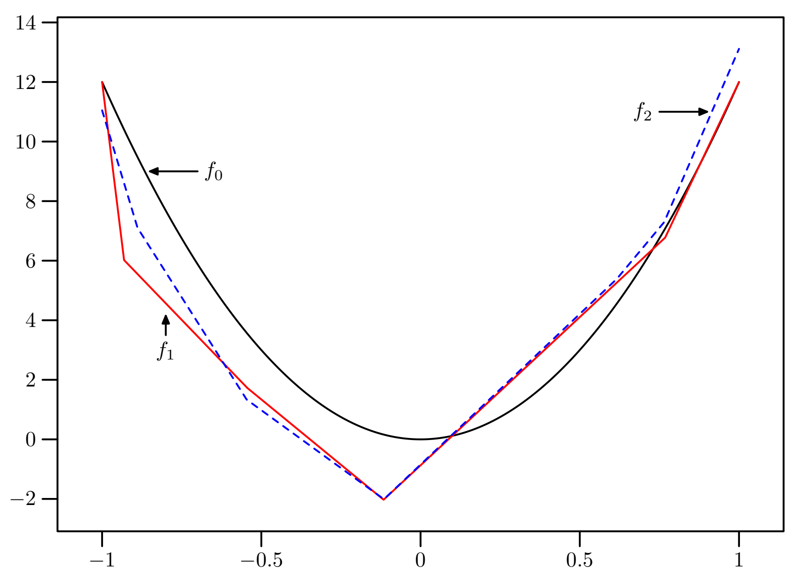

Fig. 3. *Solid: f_0 and f_1, dashed: f_2.*

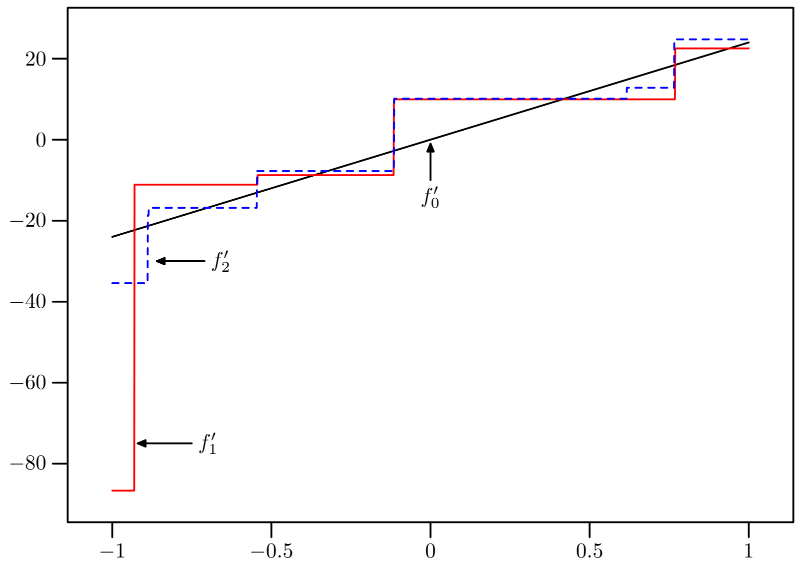

FIG. 4. *Solid:* f_0' *and* f_1', *dashed:* f_2'.

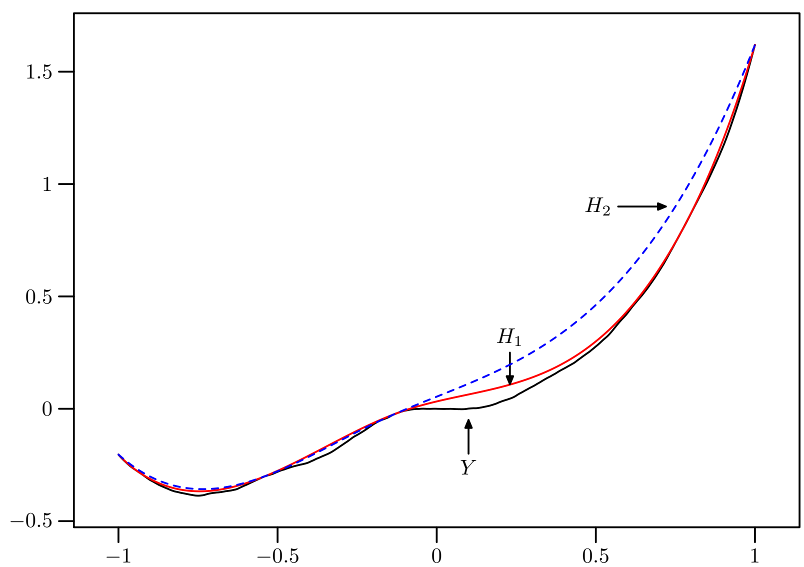

FIG. 5. *Solid:* Y *and* H_1, *dashed:* H_2. *Boundary conditions:* $f_1(\pm 1) = 12$, $f_2(\pm 1) = 6$.

4. Concluding remarks and open problems In section 2 a function of integrated Brownian motion, determining the limit distribution of nonparametric least squares estimators and maximum likelihood estimators of a convex regression function, resp. convex density, was determined. This function was called the "invelope" and uniquely characterized in Theorem 1. However, several open problems remain which we list below.

1. In the case of the limit distribution of nonparametric least squares estimators and maximum likelihood estimators of a *monotone* regression function, resp. *monotone* density, the distribution of the limit function of drifting Brownian motion was analytically characterized in GROENEBOOM (1988). In fact, the infinitesimal generator of the jump process of locations of points of touch between Brownian motion + a parabola and its convex minorant was determined analytically using Airy functions. We have no such analytic representation in the present case, and do not even know if the distribution of $H''(0)$ has finite moments.

2. We conjecture that the locations of points of touch between integrated Brownian motion $+\,t^4$ and its "invelope" are realizations of a locally finite point process ("the point are isolated"), but we have no proof. In the "monotone case" the locations of points of touch between Brownian motion + a parabola and its convex minorant are indeed realizations of a locally finite point process, but we get this from the analytic characterization of the point process in GROENEBOOM (1988).

3. Assuming that the locations of points of touch between integrated Brownian motion $+\,t^4$ and its "invelope" are realizations of a locally finite point process, will the locations of changes of slope of the solutions f_c of the constrained

minimization problem of Lemma 2.2 stay fixed in a finite interval, say $[-1,1]$, for all values of $c \geq c_0$, where c_0 may depend on the sample path of integrated Brownian motion, or will they continue to change, as $c \to \infty$? The spline relation (3.2) in section 3 suggests that they *must* continue to change, unless f_c does not change at these points either.

4. For the "monotone case" it was shown in Groeneboom (1988) that between points of touch of Brownian motion + a parabola and its convex minorant, Brownian motion behaves as "an excursion above a parabola". We conjecture that similarly, between points of touch of integrated Brownian motion $+\,t^4$ and its "invelope", integrated Brownian motion behaves as an excursion below a cubic polynomial and has a behavior that can be described with the help of the theory, developed in Groeneboom, Jongbloed & Wellner (1999). But a first step in this direction refers us back to the unsolved problem mentioned in point 2, i.e., proving that the points of touch are indeed isolated.

5. It would be of interest to consider the following "continuous time" or "white noise" regression problem, Suppose we observe $\{X(t):\ t \in [-c,c]\}$ where, for a two-sided Brownian motion W,

$$dX(t) = f(t)dt + \sigma dW(t),$$

and the "regression function" or "signal" f is assumed to be convex. Then our "canonical" convex function $t \mapsto 12t^2$ is replaced by a more general convex function f. We conjecture that the theory, developed in section 2, can be used again, and that, in particular, one gets a similar asymptotic behavior of the solutions f_c on a bounded interval $[-1,1]$, as $c \to \infty$, if the underlying regression function is strictly convex, where the limiting behavior is again described by an "invelope" of integrated Brownian + the second integral of the convex function.

6. If, in the finite sample situation, the restriction that f' is monotone is replaced by the restriction that $f^{(k)}$ is monotone, where $k > 1$, we think that the asymptotic behavior of the solution will involve a function of iteratively integrated Brownian motion, but the theory for this situation still has to be developed.

5. Appendix: Gaussian scaling relations. Suppose that for $a, \sigma > 0$ and $t \in \mathbb{R}$ we define

$$Y_{a,\sigma}(t) \equiv at^4 + \sigma \int_0^t W(s)\, ds$$

where W is standard two-sided Brownian motion. We take $Y_{1,1} \equiv Y$ to be the standard (or canonical) version of the family of processes $\{Y_{a,\sigma} : a > 0, \sigma > 0\}$. Let $H_{a,\sigma}$ be the invelope process corresponding to the process $Y_{a,\sigma}$.

Proposition 1. (Scaling of the processes $Y_{a,\sigma}$ and the invelope processes $H_{a,\sigma}$.)

$$Y_{a,\sigma}(t) \stackrel{\mathcal{D}}{=} \sigma(\sigma/a)^{3/5} Y((a/\sigma)^{2/5} t) \tag{5.1}$$

as processes for $t \in \mathbb{R}$, and hence also

$$H_{a,\sigma}(t) \stackrel{\mathcal{D}}{=} \sigma(\sigma/a)^{3/5} H((a/\sigma)^{2/5} t) \tag{5.2}$$

as processes for $t \in \mathbb{R}$.

Corollary 5. *For the invelope processes at 0 it follows that*

$$(H''_{a,\sigma}(0), H'''_{a,\sigma}(0)) \stackrel{\mathcal{D}}{=} (\sigma^{4/5} a^{1/5} H''(0), \sigma^{2/5} a^{3/5} H'''(0))\,. \tag{5.3}$$

COROLLARY 6. (Finite interval scaling.)

$$\sigma^{-8/5}a^{3/5}Y_{a,\sigma}((\sigma/a)^{2/5}t) \stackrel{\mathcal{D}}{=} Y(t), \qquad t \in [-c, c]\,, \tag{5.4}$$

and hence observation of $\{Y(t) : t \in [-c, c]\}$ *is equivalent to observation of* $\{Y_{a,\sigma}(t) : t \in [-1, 1]\}$*, if* $c = (a/\sigma)^{2/5}$.

Remark: Note that this makes some intuitive sense; σ represents the "noise level" or standard deviation of the noise and the variance of our "estimators" $H^{(k)}_{a,\sigma}(0)$, $k = 2, 3$, should converge to zero as $\sigma \to 0$. Similarly, $a =$ some constant times the curvature of the function $12at^2$ at zero; the function gets easier to estimate at this point as the curvature goes to zero, and the proposition makes this precise. Note that the scaling in (5.3) is consistent with the finite-sample convergence results of GROENEBOOM, JONGBLOED AND WELLNER (2001A) with the identification $\sigma = n^{-1/2}$.

Proofs. Starting with the proof of Proposition 5.1, we will find constants k_1, k_2 so that

$$k_1 Y_{a,\sigma}(k_2 t) \stackrel{\mathcal{D}}{=} Y(t)\,. \tag{5.5}$$

Since $\alpha^{-1/2}W(\alpha u) \stackrel{\mathcal{D}}{=} W(u)$ for each $\alpha > 0$,

$$\begin{aligned} Y_{a,\sigma}(t) &\stackrel{\mathcal{D}}{=} at^4 + \sigma\alpha^{-1/2}\int_0^t W(\alpha s)ds \\ &= at^4 + \sigma\alpha^{-3/2}\int_0^{\alpha t} W(u)du \end{aligned} \tag{5.6}$$

by changing variables. Now by (5.6)

$$k_1 Y_{a,\sigma}(k_2 t) \stackrel{\mathcal{D}}{=} k_1 a(k_2 t)^4 + k_1\sigma\alpha^{-3/2}\int_0^{k_2\alpha t} W(s)ds \tag{5.7}$$

$$= t^4 + \int_0^t W(u)du \tag{5.8}$$

if we choose k_1, k_2, α so that

$$ak_1k_2^4 = 1\,, \qquad \alpha k_2 = 1\,, \qquad \text{and} \qquad \sigma\alpha^{-3/2}k_1 = 1\,. \tag{5.9}$$

This yields $\alpha = 1/k_2$, and hence (from the last equality in the last display)

$$\sigma k_1 k_2^{3/2} = 1\,.$$

This in turn implies that

$$\frac{a}{\sigma} k_2^{5/2} = 1 \quad \text{or} \quad k_2 = (\sigma/a)^{2/5}\,.$$

This yields $k_1 = (1/\sigma)(a/\sigma)^{3/5}$. Expressing (5.5) as

$$Y_{a,\sigma}(k_2 t) \stackrel{\mathcal{D}}{=} k_1^{-1} Y(t/k_2)$$

with $k_1^{-1} = \sigma(\sigma/a)^{3/5}$ and $1/k_2 = (a/\sigma)^{2/5}$ yields the first claim of the proposition. The second claim follows from immediately from (5.2) and the definitions of $H_{a,\sigma}$ and H.

Corollary 5 follows from (5.3) and straightforward differentiation.

To prove Corollary 6, note that (5.2) is equivalent to

$$\sigma^{-8/5} a^{3/5} Y_{a,\sigma}(t) \stackrel{\mathcal{D}}{=} Y(t)\,.$$

Hence observation of Y on the interval $[-c, c]$ is equivalent to observation of $\sigma^{-8/5} a^{3/5} Y_{a,\sigma}(t)$ for $t \in [-1, 1]$ if $c = (a/\sigma)^{2/5}$. □

Acknowledgments. We owe thanks to: Enno Mammen for several interesting conversations about his work on convex regression problems; Mary Meyer and Yazhen Wang for sending us copies of unpublished Ph.D. dissertation work and subsequent technical reports; and Peter van der Wijden for help in producing Figures 1 to 5 in section 3. We also wish to thank two anonymous referees and the associate editor for their remarks which led to improvements in the presentation of the paper.

REFERENCES

Brunk, H.D. (1970). Estimation of isotonic regression. *Nonparametric Techniques in Statistical Inference* Puri, M.L. ed.

Dieudonné, J. (1969). *Foundations of Modern Analysis*. Academic Press, New York.

Groeneboom, P. (1985). Estimating a monotone density. *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer,* Vol. II, Lucien M. LeCam and Richard A. Olshen eds.

Groeneboom, P. (1988). Brownian motion with a parabolic drift and Airy functions. *Probability Theory and Related Fields* **81**, 79 - 109.

Groeneboom, P. and Wellner, J. A. (2000). Computing Chernoff's distribution. *J. Computational and Graphical Statistics*, to appear.

Groeneboom, P., Jongbloed, G., and Wellner, J.A. (1999). Integrated Brownian motion conditioned to be positive. *Ann. Probability* **27**, 1283-1303.

Groeneboom, P., Jongbloed, G., and Wellner, J.A. (2001a). Estimation of convex functions: characterizations and asymptotic theory. Submitted.

Groeneboom, P., Jongbloed, G., and Wellner, J. A. (2001b). Computation of nonparametric function estimators via vertex direction algorithms. Manuscript in progress.

Jongbloed, G. (1995). *Three Statistical Inverse Problems.* Unpublished Ph.D. dissertation, Delft University of Technology, Delft.

Mammen, E. (1991). Nonparametric regression under qualitative smoothness assumptions. *Ann. Statist.* **19**, 741-759.

Meyer, M.C. (1997). *Shape Restricted Inference with Applications to Nonparametric Regression, Smooth Nonparametric Function estimation, and Density Estimation.* Ph.D. Dissertation, Department of Statistics, University of Michigan, Ann Arbor, Michigan.

Prakasa Rao, B.L.S. (1969). Estimation of a unimodal density. *Sankhȳa.* Ser. A, **31**, 23-36.

Press, W.H., Teukolsky, S.A., Vetterling, W.T. and Flannery, B.P. (1992). *Numerical Recipes in C.* Cambridge University Press, Cambridge.

Rogers, L.C.G., and Williams, D. (1994). *Diffusions, Markov Processes and Martingales, Volume 1, 2nd edition.* Wiley, New York.

Sinai, Y. G. (1992). Statistics of shocks in solutions of inviscid Burgers equation. *Comm. Math. Phys.* **148**, 601 - 621.

Wang, Y. (1994). The limiting distribution in concave regression. *Preprint*, University of Missouri, Columbia.

Wright, S. J. (1997). *Primal-Dual Interior Point Methods*. SIAM, Philadelphia.

Department of Mathematics
Delft University of Technology
P.O. Box 356
2600 AJ Delft
The Netherlands
e-mail: P.Groeneboom@its.tudelft.nl

Department of Mathematics
Vrije Universiteit
1081 HV Amsterdam
The Netherlands
e-mail: geurt@cs.vu.nl

University of Washington
Statistics
Box 354322
Seattle, Washington 98195-4322
U.S.A.
e-mail: jaw@stat.washington.edu