Graduated Non-Convexity by Smoothness Focusing

Mads Nielsen
DIKU,
University of Copenhagen,
Universitetsparken 1, DK-2100 Ø

Abstract

Noise-corrupted signals and images can be reconstructed by regularization. If discontinuities must be preserved in the reconstruction, a non-convex solution space is implied. The solution of minimum energy can be approximated by the Graduated Non-Convexity (GNC) algorithm. The GNC approximates the non-convex solution space by a convex solution space, and varies the solution space slowly towards the non-convex solution space. This work provides a method of finding the convex approximation to the solution space, and the convergent series of solution spaces. The same methodology can be used on a wide range of regularization schemes. The approximation of the solution space is carried out by a scale space extension of the smoothness measure, in which a coarse-to-fine analysis can be performed. It is proven, that this scale space extension yields a convex solution space. GNC by smoothness focusing is tested against the Blake-Zisserman formulation and is shown to yield better results in most cases. Furthermore, is it pointed out that Mean Field Annealing (MFA) of the weak string does not necessarily imply GNC, but behaves in a predictable and inexpedient manner.

1 Introduction

Regularization is a method of reformulating ill-posed inverse problems as well-posed problems as done by Tikhonov and Arsenin [1]. This reformulation implies the addition of a stabilizing term, followed by a global minimization in a convex solution space, yielding a unique solution. As computer vision can be regarded as inverse optics, many computer vision problems are ill-posed by nature as argued by Aloimonos and Shulman [2]. Therefore, regularization is often applied in computer vision. By standard regularization of the surface reconstruction problem, discontinuities will be smoothed. As discontinuities plays an important role in vision, schemes for discontinuity preserving regularization has been proposed by eg. Geman and Geman [3]. The schemes of discontinuity preserving regularization implies minimization in a non-convex solution space, why the avoidance of local minima in the optimization procedure becomes a problem. Optimization in the non-convex solution space has been performed by simulated annealing by Geman and Geman [3], by genetic algorithms by Jensen and Nielsen [6] and by the deterministic approaches of Graduated Non-Convexity (GNC) by Blake and Zisserman [7] and Mean Field Annealing by Geiger and Girosi [14].

Blake and Zisserman have developed the deterministic and approximative approach of GNC [7], which implies an approximation of the non-convex solution space by a convex solution space. This approximation of the solution space is slowly varied towards the non-convex solution space, in the hope that the local

BMVC 1993 doi:10.5244/C.7.60

minimum, which is tracked, will converge to the global minimum of the non-convex solution space. Blake and Zisserman have found an approximation of the solution space in special cases, such as the weak string. Blake has shown, that in general the GNC is faster than simulated annealing [8], why GNC is desirable in many other situations than the weak string. This work presents a general scheme for creating GNC algorithms. It is proven, that any stabilizing term, which is determined as the sum of functions of the local derivatives of the solution, can be approximated by a Gaussian scale space extension, yielding a convex solution space (The adiabatic approximation used by Rangarajan and Chellappa [9] can be used to put non-local derivative-interactions on a local form). By slowly varying the standard deviation of the Gaussian towards zero, the solution space of the approximation will slowly vary towards the non-convex solution space. The scheme corresponds to the scheme in any coarse-to-fine analysis as eg. the edge focusing by Bergholm [10]: Make a scale space extension, detect the solution on the highest scale and track the solution to lower scales to gain precision in the solution.

2 Surface reconstruction

An example of an ill-posed problem is the reconstruction of a signal s corrupted with stationary and additive noise n yielding the measured signal c. The solution $\tilde{s}$ to the reconstruction problem can be found by standard regularization as minimization of $E(\tilde{s}) = \int (c - \tilde{s})^2 + \lambda \tilde{s}_x^2 dx$ where subscript denotes the derivative and λ is a weighing between smoothness and data term. The term, which is to be minimized is called the energy, the first term in the integral is called the data term and the second term is called the stabilizing or smoothness term.

If the underlying function s contains discontinuities, and these must be preserved during the regularization, schemes like the line process by Geman and Geman [3] or the equivalent smoothness thresholding by Blake and Zissermann [7] can be used. The elimination of the line process yielding the smoothness thresholding is a special case of the adiabatic approximation [9]. The idea is to punish derivatives no more than some certain value even if it grows towards infinity. In this case the solution is not influenced by any force from the smoothness term, when the derivative is high, and the solution can totally adapt to the data, and thereby preserve the discontinuities. The two different formulations of the discrete discontinuous regularization is given in Equation 1 (line process) and Equation 2 (smoothness thresholding).

$$E(\tilde{s}) = \sum_{\tilde{s}} (c - \tilde{s})^2 + \lambda(\eta \tilde{s}_x^2 + (1 - \eta) T^2) \tag{1}$$

where E has to be minimized over $\eta(x)$ as well as $\tilde{s}(x)$, and T^2 is a constant punishment for detection of discontinuities. In the minimized solution, η will always yield 1 or 0, dependent on whether $\tilde{s}_x^2 > T^2$ or not. The smoothness thresholding formulation is

$$E(\tilde{s}) = \sum_{\tilde{s}} (c - \tilde{s})^2 + f(\tilde{s}_x) \qquad \text{where } f(t) = \lambda \begin{cases} t^2 & \text{if } t^2 < T^2 \\ T^2 & \text{otherwise} \end{cases} \tag{2}$$

where T is the same constant as in Equation 1. The solution, which minimizes the energy of Equation 1 and Equation 2 will be piecewise continuous, and then in a finite number of points be discontinuous. Whether a point yields a discontinuity depends on the input signal, the discontinuity threshold T and the weighing λ. The interaction between the factors is not simple. For a discussion see Nielsen [4].

Many other smoothness functions f has been proposed. In general, the smoothness functions are quadratic in the derivative for values close to zero. If discontinuities are to be preserved, the smoothness function is made less increasing for larger values of the derivative. The dependency on larger values of the derivative might be a constant function, a linear function, a logarithmic function, or any other less than quadratic increasing function (see Nielsen [5]).

The GNC algorithm works by changing the smoothness function f to create the convex solution space. In the following an example of the GNC approximation of the solution space is outlined, before discussing the implications and other general methods.

3 GNC of the weak string

The weak string is defined by minimization of the energy in Equation 2. The solution space is not convex. Therefore, a stochastic optimization technique or deterministic approximation has to be applied. The GNC is a deterministic approximation.

In Appendix A Lemma 1, it is shown, that the solution space is convex if the second derivative of the smoothness function f in Equation 2 is bounded to be larger than $-\frac{1}{2}$. This is obviously not the case in Equation 2 where the second derivative does not even exist in $t^2 = T^2$. The idea of the GNC algorithm is to approximate f by another function, which is limited in the second derivative. Blake and Zisserman obtain this by approximating f by f_1 in the region where $t^2 \approx T^2$ by a second order polynomial with a second derivative larger than $-\frac{1}{2}$.

If we denote the initial smoothness function f by f_0, we can construct a series of functions f_c which is continuously varying as a function of c. When $c = 1$ the solution space is convex, and when $c = 0$ the regularization corresponds to the weak string. The intermediate functions f_c, $c \in]0;1[$ can be constructed by letting the interval of approximation shrink to a factor c of the original interval. The claim of Blake and Zisserman is that if we track the local minimum of f_c, when slowly varying c from 1 to 0, we will obtain a good approximation to the solution of the weak string [7]. It is shown by Blake and Zisserman [7], that the global minimum cannot always be tracked as the local minimum. A discussion of convergence is given by March [13] and Nielsen [11].

Other approximations to the initial smoothness function, which yield a convex solution space, can be constructed. In the following the scale space extension is proposed. In Appendix A it is proven that the weak string has a convex solution space if the smoothness function is convoluted by a Gaussian of appropriate standard deviation. In fact, it is proven that any smoothness function, which only differs from the one of convex solution space by a Lebesgue integrable function, will cause a convex solution space if the smoothness function is filtered with a Gaussian of appropriate standard deviation. In this way a convex solution space can be constructed, and the solution can be found using a simple gradient descend algorithm. When slowly decreasing the standard deviation of the Gaussian towards zero, we can track the solution to the optimization problem by tracking the minimum as a local minimum. When the standard deviation yields zero, the tracked solution is an approximation of the solution to the original problem.

This is a general scheme of constructing a GNC algorithm, which only requires that the smoothness function fulfill the demands of Lemma 2 in Appendix A. The algorithm is presented as follows:

$\sigma = \sigma_0$
while $\sigma > \sigma_1$
  Minimize $E_\sigma(\tilde{s}) = \sum (c - \tilde{s})^2 + f_\sigma(\tilde{s}_x)$
  $\sigma = \sigma / \delta$
endwhile

where σ is the standard deviation of the Gaussian used for convolution of the smoothness function f. In Appendix A a conservative measure of σ_0 is given. In practice we might find σ_0 by iteratively increasing σ and test the second derivative.

The smoothness focusing can be interpreted from a view of Bayesian estimation. If the measurements of the derivatives of the intermediate solutions are perceived as noisy, with Gaussian uncorrelated noise, the smoothness term should as an approximation be convoluted by a Gaussian to yield a Maximum A Posteriori (MAP) estimate. For a more thorough explanation see Nielsen [12].

4 Mean Field Annealing and GNC

Mean Field Annealing (MFA) is a technique, which is a deterministic version of Simulated Annealing. Instead of simulating the stochastic behaviour of a molecule, the mean state of all possible states of the molecule is simulated. This implies, that the MFA is deterministic. To find the mean state one has to integrate the probability of all possible states as defined by the *partition function* [15]. The partition function is not always easily integrable, and approximations has to be carried out very carefully. Geiger and Girosi claims, that the MFA of the weak string is a GNC algorithm. This is not the case as the solution space might always be non-convex. In the work of Geiger and Girosi [14], the smoothness function has the form:

$$f_\sigma(x) = T^2 - \sigma \log(1 + e^{\frac{(T^2 - \lambda x^2)}{\sigma}})$$

where σ can be interpreted as the temperature. This has the characteristics of being a smoothed version of the weak string. When σ approaches zero, the smoothness function approaches the weak string. When σ approaches infinity, the lower bound on the second derivative approaches a negative limit, which is dependent on λ. The limit is empirically found to be approximately 0.6λ, which shows, that the MFA is not a GNC for $\lambda > 0.9$ approximately. The positions on the smoothness function, where the second derivative takes its minimum value is approximately a linear function of the temperature, for high temperatures. This means that the concavities in the solution space is placed in $\pm k\sigma$, where k is some constant. If σ is initialized to a value, such that every derivative in the signal is in the interval $[-k\sigma; k\sigma]$, no discontinuities will be detected using this σ. As σ is decreased slowly, the concavities traverse towards zero. As we track the solution as a local minimum, the solution stays in the interval, when σ is lowered. This means that no discontinuities are detected if the temperature is started high enough, because the derivatives are pulled towards zero. The consequence of the MFA is, that the initial temperature defines the discontinuity positions. Implementations of Geiger and Girosis MFA algorithm [14], where the line process has been eliminated as done by Blake and Zisserman [7], and implementations of the MFA of the line process directly as proposed by Hansen [16] yield identical results and both show this tendency very clearly.

The argumentation of using MFA is solely based upon statistical physics. In statistical physics Mean Field Theory is not regarded as a good approximation inside the critical regions. The weak string is normally situated in a critical region. A critical region is the regions of the parameter settings, where phase transitions are present. If both discontinuity points and non-discontinuity points are present, the weak string is in the critical region. If not, another model than the weak string could have been used.

The Generalized Graduated Non-Convexity algorithm [9] gives a method of eliminating non-local interactions by adiabatic approximation. It does, however, not in general guarantee a GNC algorithm. The GGNC is a broad class of algorithms, among which some are GNCs. The scale space extension is a special and constructive case of the GGNC.

5 Experiments

Whether the approximation using the scale space extension is better or worse than special designed GNC approximations [7] is not easy to judge. The specially designed approximation has the advance, that the smoothness function is only changed around the critical points where the second derivative is smaller than $-\frac{1}{2}$. The Gaussian convolution yields the theoretically satisfying property of being explainable from probability theory [12]. In the following the resemblance and difference is analysed. In [5] the Gauss GNC is tested on 2D images.

The outcome of the GNC algorithm implemented as done by Blake and Zisserman [7] is compared to the Gaussian convolution of the smoothness function. The testproblem is the minimization of the energy in Equation 2. The Blake and Zisserman approximation is outlined in Section 3. The scale space extension of the smoothness function is:

$$\begin{aligned} f_\sigma(x) &= T^2 + \frac{\sigma^2 + x^2 - T^2}{2}(\operatorname{erf}(x_-) + \operatorname{erf}(x_+)) \\ &- \frac{x\sigma}{\sqrt{2\pi}}(e^{-x_-^2} - e^{-x_+^2}) - \frac{T\sigma}{\sqrt{2\pi}}(e^{-x_-^2} + e^{-x_+^2}) \end{aligned}$$

where

$$x_- = \frac{\sqrt{2}(x-T)}{2\sigma} \qquad x_+ = \frac{\sqrt{2}(x+T)}{2\sigma} \quad \text{and} \quad \operatorname{erf}(x) = \frac{2}{\sqrt{\pi}}\int_{-\infty}^{x} e^{-t^2}dt$$

The qualitative difference between the Blake-Zisserman approximation and the scale space approximation is that the scale space not only rounds off the corners, it also increases the value in zero. In Figure 1 the two different approximations are shown. In the scale space extension, the second derivative only yields the critical value of $-\frac{1}{2}$ in two points. The Blake-Zisserman approximation yields the critical value in two rather wide intervals. In these intervals, the solution space is nearly

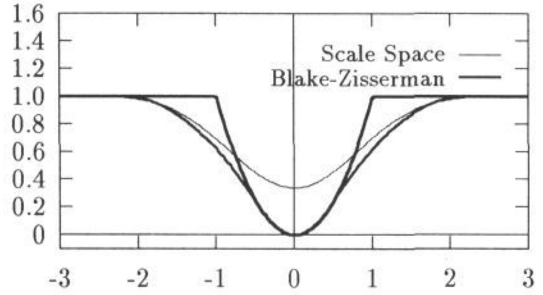

Figure 1: Smoothness function of the derivative of the solution in starting level of GNC as formulated by Blake and Zisserman and in scale space extension.

non-convex, and might have a zero-gradient. This means that a gradient descend algorithm will probably end up in one of the ends of the interval at random. Whether a gradient in the input is perceived as a discontinuity or not, might in this way be random. Two types of experiments have been performed. One, on a noise-corrupted signal, and one on an ideal and precisely adjusted signal. The latter to test the precise behaviour to certain features, the first to make an overall judgment.

An ideal signal consisting of an interval of negative gradient, an interval of zero gradient, a step edge, and an interval of zero gradient has been noise corrupted with stationary Gaussian noise with standard deviation $\sigma = 2.5$. The two GNC algorithms has been run on the signal. The result of the Gauss GNC can be seen in Figure 2. The Blake-Zisserman GNC detects two more discontinuities (in the high gradient interval) than the Gauss GNC. The Gauss GNC yields a final energy which is 88 percent of the total energy of the energy found by the Blake-Zisserman algorithm. This is a general tendency which is emphasized in the next experiment.

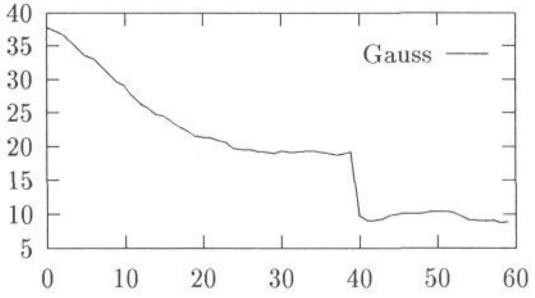

Figure 2: Regularized signal using the weak string approximation by Gauss GNC.

The two GNC algorithms are not always detecting the same discontinuities. It is well know [7], that the weak string will detect discontinuities from a gradient, if the gradient $g > \frac{T}{\sqrt{2\lambda}}$ [4]. In this experiment, the algorithms has been tested on a constant gradient. From each experiment to the next, the gradient has been increased. In Figure 3 the energy of the solution found by the two algorithms is plotted as a function of the gradient. In regions, where the solution is not changing the detection of discontinuities, the plot should be a parabola. For each combination of discontinuities a parabola exists. The perfect GNC algorithm would for a given gradient choose the parabola of lowest energy. This is not the case for any of the two algorithms evaluated in this paper.

The experiment shows, that initially, where the gradient is small, non of the algorithms detect any discontinuities, and the solutions are thereby identical. From a certain point (around $g = 0.6$) the Blake-Zisserman GNC detects discontinuities, and thereby leaves the initial parabola. The energy increases relative to the initial parabola, and it can be concluded that the discontinuities has been detected too early. In Figure 4 a zoom-in on the region of differences can be seen. The Gauss GNC follows the initial parabola until the gradient g is close to 0.8. After this it follows a new parabola. The energy drops from the first to the second parabola, and it can be concluded, that discontinuities has been detected too late. In the region of larger gradients (approximately 1.6), the Gauss GNC result in the worst solution, as too many discontinuities has been detected. This region is, though, of less interest, because it is the region where nearly all points has been detected as discontinuities. This situation should never appear in a realistic environment.

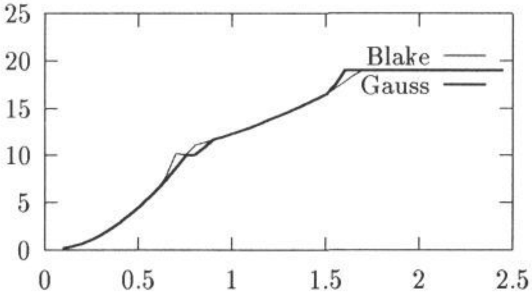

Figure 3: The energy of the weak string as a function of the gradient. For each gradient a signal of length 20 and constant gradient has been constructed. The energy is plotted for the Blake-Zisserman GNC and for the Gauss GNC. $\lambda = T = 1.0$ in all computations. In the interval of gradients $0.6 < g < 0.95$ the Blake-Zisserman GNC yields a higher energy than the Gauss GNC except in the region $0.75 - 0.78$.

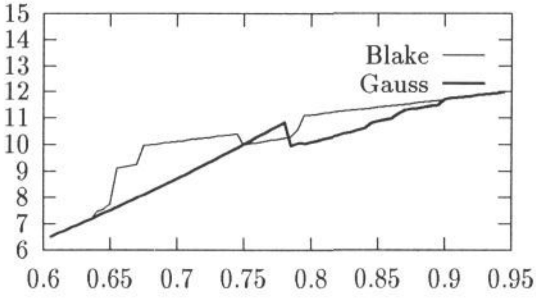

Figure 4: The energy of the weak string as a function of the gradient. Computations performed as in 3. Focus is on the region of different energy.

6 Conclusion

A general method to construction of Graduated-Non-Convexity algorithms has been proposed. The method (called Gauss GNC) can be used for approximation of any penalty function, which is a function of a single derivative of the solution. It implies a scale space extension of the penalty function. It is proven, that the method yields convex approximations for any penalty function, which can be constructed as the sum of a smoothness function with convex solution space and a Lebesgue integrable function. The method can be derived from estimation theory.

In the case of the weak string, the application is straight-forward, and yields results, which in general are better than those of Blake and Zisserman. The Blake-Zisserman GNC has a tendency to over-estimate the number of discontinuities, while the Gauss GNC has a tendency to under-estimate the number of discontinuities. It is shown, that the tendency in general is the worst for the Blake-Zisserman GNC.

Earlier, Mean Field Annealing has been used to make deterministic approximations to the process of simulated annealing of the weak string. MFAs does not yield a GNC algorithm of the weak string, as the solution space not is convex even for infinite high temperatures. The start temperature of the MFA defines the positions of the discontinuities. No matter how low the discontinuity threshold is, it can be matched by a starting temperature, which results in no discontinuities.

The Gauss GNC implies the possibility of automatically applying GNC to any regularization, where the smoothness function is function of only one derivative of the solution. In a general situation, an analytic expression of the penalty function is not needed. This implies the possibility of using a penalty function, which is measured as a histogram, and then only numerically known. In this way a new category of GNC applications is made possible.

I want to acknowledge M. B. Larsen, S. I. Olsen for discussions of GNC, P. Johansen and K. Grue for discussions of scale space theory, and Jens Jensen for discussion of Mean Field Theory.

References

[1] A. N. Tikhonov and V. Y. Arseninn "*Solution of ill-posed problems*" V.H Winston & Sons, John Wiley, 1977.

[2] J. Aloimonos and D. Shulman, "*Integrating of visual modules*" Academic Press, 1989.

[3] S. Geman and D. Geman, "*Stochastic Relaxation, Gibbs Distribution, and the Bayesian Restoration of Images*", IEEE PAMI Vol. 6, No. 6, pp. 721–741, April 1984.

[4] M. Nielsen, "*Scale Space Extensions of the Weak String*", Proc. 8th SCIA, Tromsø, Norway, May 25-28, 1993.

[5] M. Nielsen, "*Isotropic Regularization*", Proc. 4th BMVC, Guildford, Surrey, September 21-23, 1993.

[6] J. B. Jensen and M. Nielsen, "*A Simple Genetic Algorithm Applied to Discontinuous Regularization*", Proc. IEEE workshop on NNSP, Copenhagen, 29/8-2/9, 1992.

[7] A. Blake and A. Zisserman, "*Visual Reconstruction*" MIT Press, 1987.

[8] A. Blake, "*Comparison of the Efficiency of Deterministic and Stochastic Algorithms for Visual Reconstruction*" Proc. of PAMI Conference, 1989.

[9] A. Rangarajan and R. Chellappa, "*Generalized Graduated Non-Convexity Algorithm for Maximum A Posteriori Image Estimation*" Proc. of 10th ICPR, Atlantic City, New Jersey, USA, 16-21/6,1990.

[10] F. Bergholm, "*Edge Focusing*", IEEE Trans. PAMI-9, No. 6, pp 726-741, November, 1987.

[11] M. Nielsen, "*Discontinuous regularization interpreted as scale space*", Project 16-6, DIKU, University of Copenhagen, 1992.

[12] M. Nielsen, "*Graduated Non-Convexity by Smoothness Focussing*' Technical report 18-5-93, DIKU, University of Copenhagen, 1993.

[13] R. March. "*Visual reconstruction with discontinuities using variational methods*" Image and Vision Computing, vol. 10, no. 1, Jan/Feb 1992.

[14] D. Geiger and F. Girosi, "*Parallel and Deterministic Algorithms from MFRs: Surface Reconstruction*", IEEE PAMI, Vol. 13, No. 5, May 1991.

[15] I. Lifnitz and I. Landau, "*Statistical Physics*", Pergammon Press, 1982.

[16] L. K. Hansen, "*Bolzman Learning of Parameters in Bayes Visual Reconstruction*" Proc. 1st Danish Conference on Pattern Recognition and Image Analysis, DIKU, Copenhagen, 11-12/6, 1992.

Appendix A

In this appendix, the proofs of convexity of the solution space, when the smoothness function is scale space extended is given. The proof falls in two lemmas. The first concerns the conditions on the smoothness function to create a convex solution space. The second concerns the scale space extension of functions which behaves nicely except for an integrable part.

Lemma 1 *If the second derivative of the smoothness function f can be bounded downwards to $-\frac{1}{2}$ the solution space will be convex if the energy E of the discrete sampled signal is given by*

$$E(\tilde{s}) = \sum (\tilde{s}_i - c_i)^2 + f(\tilde{s}_i - \tilde{s}_{i-1}) \tag{3}$$

Proof The solution space is convex if the *Hessian Matrix* $\mathcal{H}$ of the energy is positive definite. In this case

$$\mathcal{H}_{ij} = \frac{\partial^2 E}{\partial \tilde{s}_i \partial \tilde{s}_j} = \begin{cases} -f_i'' & \text{if } i = j+1 \\ 2 + f_i'' + f_{i+1}'' & \text{if } i = j \\ -f_j'' & \text{if } j = i+1 \end{cases}$$

$$\text{where } f_i'' = \begin{cases} f''(\tilde{s}_i - \tilde{s}_{i-1}) & \text{if } 2 \leq i \leq N \\ 0 & \text{otherwise} \end{cases}$$

By definition of positive definiteness the solution space is convex if:

$$\forall \vec{x} : \sum_1^N 2x_i^2 + \sum_2^N (x_i - x_{i-1})^2 f_i'' > 0$$

Because $\sum (2x_i)^2 \geq \sum (x_i - x_{i-1})^2$ for all $\vec{x}$, $\mathcal{H}$ is positive definite if $\forall i : f_i'' > -\frac{1}{2}$
End of proof

The lower bound on the second derivative can be reached by a convolution of the smoothness function f by a Gaussian with an adequate standard deviation σ if the smoothness function only differs from the convex smoothness function by a Lebesgue integrable function.

Lemma 2 *If we let b be any constant, $*$ denote the convolution, and $G(x,\sigma)$ be the Gaussian in x of standard deviation σ we have for any function f which*

$$\forall\epsilon > 0 : \forall x \in I\!R : f(x) = g(x) + h(x),\ \frac{\partial^2}{\partial x^2} g(x) \geq -b^2 + \epsilon,\ \int_{I\!R} |h(x)| dx = A$$

implies that

$$\sigma > \left(\frac{A\sqrt{2\pi}e^{-3/2}}{2\epsilon}\right)^{1/3} \Rightarrow \forall x \in I\!R : \frac{\partial^2}{\partial x^2}(f(x) * G(x,\sigma)) > -b^2$$

Proof We have

$$\begin{aligned}
\frac{\partial^2}{\partial x^2}(G * f) &= \frac{\partial^2}{\partial x^2}(G * g) + \frac{\partial^2}{\partial x^2}(G * h) \\
&= \int_{\mathrm{IR}} G(k,\sigma) \frac{\partial^2}{\partial x^2} g(x-k) dk + \int_{\mathrm{IR}} h(k) \frac{\partial^2}{\partial x^2} G(x-k,\sigma) dk \\
&\geq \int_{\mathrm{IR}} (-b^2+\epsilon) G(k,\sigma) dk - \int_{\mathrm{IR}} |h(x)| |\frac{\partial^2}{\partial x^2} G(x-k,\sigma)| dk \\
&\geq (\epsilon - b^2) \int_{\mathrm{IR}} G(k,\sigma) dk - \int_{\mathrm{IR}} |h(k)| \sup_{i \in \mathrm{IR}} |\frac{\partial^2}{\partial x^2} G(x-i,\sigma)| dk \\
&= -b^2 + \epsilon - \sup_{i\in \mathrm{IR}} |\frac{\partial^2}{\partial x^2} G(x-i,\sigma)| \int_{\mathrm{IR}} |h(k)| dk \\
&= -b^2 + \epsilon - \frac{2e^{-3/2}}{\sqrt{2\pi}\sigma^3} A
\end{aligned}$$

This is a lower bound on the second derivative of the convolution. This bound should be greater than $-b^2$ to prove the lemma.

$$-b^2 + \epsilon - \frac{2e^{-3/2}}{\sqrt{2\pi}\sigma^3} A > -b^2 \quad \Leftrightarrow \quad \sigma > \left(\frac{A\sqrt{2\pi}e^{-3/2}}{2\epsilon}\right)^{1/3} \tag{4}$$

End of proof

As an example, we can mention the weak string. The smoothness function of the weak string can be described as a constant function, plus a negative parabola in a limited region. By simple calculations we find, that the solution space is guaranteed convex by Eg. 4 if

$$\sigma > \left(\frac{2\lambda\sqrt{2\pi}e^{-3/2}}{3b^2}\right)^{1/3} T \approx 0.91 \lambda^{1/3} T$$

It should be mentioned, that this bound is a conservative measure, and in practice much smaller values of σ might yield a convex solution space. Actually, in practice we find the limit for the weak string of first order regularization to be 30% lower in the example used earlier in this paper.