Visual cortex mapping by conjugate gradients

D. H. Smith[1]

(Received 26 July 2008; revised 2 October 2008)

Abstract

In the absence of ocular dominance, conjugate gradient iteration applied to a certain objective function formulation reproduces the characteristic orientation preference map structure previously derived from a related dynamical system approach on a coarser cortex mesh. Numerical experiments indicate a strong sensitivity to the line search parameters and direction update, demanding careful consideration.

Contents

http://anziamj.austms.org.au/ojs/index.php/ANZIAMJ/article/view/1373 gives this article, © Austral. Mathematical Soc. 2008. Published October 16, 2008. ISSN 1446-8735. (Print two pages per sheet of paper.)

1 Introduction

Visual cortex mapping calculations seek a set of receptive field quantities residing on a two dimensional cortex mesh [4], given a set of corresponding visual stimulus points. In one sense, the aim is to match receptive fields with stimuli, allowing the latter to gain representation in the cortex [10], while simultaneously there is a requirement to maintain proximity between the receptive fields of neighbouring cortical points, ensuring a degree of local continuity, or smoothness [9]. One way of addressing these conflicting demands is via a self organising feature map approach [10], in which the stimulus points are treated one by one. The receptive field of that cortical point which best matches a given stimulus is adjusted towards the stimulus, together with those of its immediate neighbours. Alternatively, a dynamical system approach [9] uses a two dimensional Laplacian term to cater for local smoothness, also related to 'cortical wiring length', in conjunction with a separate term based on cortical activity patterns in response to stimuli, averaged over the stimulus set. Analogous terms appear in a related objective function minimisation formulation [3], as outlined in Section 2, which provides the basis for this particular study.

Orientation preference is a fundamental receptive field quantity, associated with edges in visual stimuli, that has received considerable attention in the experimental [2] and numerical domains [4]. In the orientation context,

a certain emphasis has been afforded to key vortex-like structures known as orientation centres, or singularities around which iso-orientation domains are arranged in a pinwheel fashion. For the case of a four dimensional receptive field comprising orientation preference and retinal position, previous calculations via a dynamical system approach produced orientation preference maps characterised by large stripe-like iso-orientation domains with sparse pinwheel density [9]. Introducing the additional quantity of ocular dominance, representing the cortical preference for input from one particular eye, at various segregation levels, also demonstrated an intimate link between this quantity and the rate of pinwheel annihilation. Wolf and Geisel [9] showed that ocular dominance serves as a suppression agent for the pinwheel annihilation process, independent of model parameters and details.

Conjugate gradient iteration applied to the related objective function formulation, with a certain amount of coaxing, will reproduce the characteristic pattern of iso-orientation domains expected in the absence of ocular dominance [9], which is graphically illustrated in Section 3. Representative iteration trajectories given in Section 3 encounter some highly oscillatory gradient behaviour, symptomatic of ill-conditioning [8], and display a strong sensitivity to the search direction update scheme. Gradient spikes in the trajectories are also shown to coincide with certain spatial evolution processes involving pinwheels in the orientation preference map.

2 The objective function formulation

For an $N \times N$ discrete cortex lattice, with receptive field vector $\mathbf{u}_j$ defined at each lattice location $j = 1, \ldots, N^2$, the objective function

$$f(\mathbf{u}; \mathbf{v}, \lambda, \beta) = \frac{1}{2}\beta \mathbf{u}^T A \mathbf{u} - \lambda \sum_i \log \sum_j \exp \frac{-(\mathbf{u}_j - \mathbf{v}_i)^T (\mathbf{u}_j - \mathbf{v}_i)}{2\lambda^2}, \qquad (1)$$

where $\mathbf{u}$ is the concatenation of all $\mathbf{u}_j$, and $\mathbf{v}_i$ is a corresponding stimulus point, represents competing demands of continuity, or smoothness, from the first term and coverage of the stimulus space, or diversity, from the second term. Both terms involve sums of squared differences, the first between receptive fields of immediate cortical neighbours via the sparse matrix A, and the second between receptive fields and corresponding stimulus points, which seek to gain a representation on the cortex [10]. Continuity improvements drive the first term of (1) towards zero from above, and coverage improvements increase the second term, to be subtracted, with an overall compromise being sought by minimisation. Parameter λ represents the receptive field size, and β is the strength of lateral interactions promoting local continuity [9].

For the case of a four dimensional receptive field comprising retinal position and orientation preference, $\mathbf{u}_j = \begin{bmatrix} x_j & y_j & r_j \cos 2\theta_j & r_j \sin 2\theta_j \end{bmatrix}^T$, in which (x_j, y_j) are the centre of the receptive field in visual space [4]. Orientation is characterised by two parameters, preference angle θ and selectivity r, referring to that orientation generating the strongest response, and to the rate of response decay with departure from the preferred value, respectively [1]. Orientation preference differences appearing in each term of (1) are evaluated via cartesian components of associated vectors with angles of twice the orientation, as is done in the combination of differential images derived from pairs of orthogonal stimuli to generate experimental orientation preference maps [1], reflecting an inherent π-periodicity [10].

Each stimulus point $\mathbf{v}_i$ in (1) contains a representative of the relevant receptive field quantities. Visual field coordinates are normal cartesian coordinates residing in the unit square [4], divided into a uniform 20×20 grid, while orientation preference coordinates reside on a circle of radius equal to the orientation selectivity [4]. Six equal angle divisions and 400 visual field points combine to make a total of 2400 stimulus points used for this study.

3 Numerical results

3.1 Conjugate gradient iteration

The task at hand is one of unconstrained nonlinear optimisation on the objective function (1). With the gradient easily available, nonlinear conjugate gradient iteration [6] is an ideal candidate method, thanks to its minimal storage requirements in the face of large scale calculations with 2^{16} variables for a 128×128 cortex lattice containing four dimensional receptive fields.

A current iterate $\mathbf{u}^{(k)}$ is updated according to $\mathbf{u}^{(k+1)} = \mathbf{u}^{(k)} + s^{(k)}\mathbf{p}^{(k)}$, where the steplength $s^{(k)}$ is chosen to approximately minimise the objective function $\phi(s) = f(\mathbf{u}^{(k)} + s\mathbf{p}^{(k)})$, evaluated along the search direction $\mathbf{p}^{(k)}$, by a line search algorithm [11]. This is based on polynomial approximations to the objective function along $\mathbf{p}^{(k)}$, always maintaining the correct values of $\phi(0)$ and $\phi'(0)$. An initial quadratic approximation, agreeing with $\phi(0), \phi'(0)$ and $\phi(1)$, is minimised to yield a fourth point from which a cubic approximation is subsequently constructed and minimised. Successive cubic approximations, through the latest two points, are minimised until the exit conditions [5, 6]

$$\phi(s) - \phi(0) \leq \alpha s \phi'(0) \quad \text{and} \quad \phi'(s) \geq \eta \phi'(0) \tag{2}$$

are met, with parameters obeying $0 < \alpha < \eta < 1$. The new search direction $\mathbf{p}^{(k+1)}$ is constructed from the current direction $\mathbf{p}^{(k)}$ and new gradient $\mathbf{g}^{(k+1)}$ according to the expression $\mathbf{p}^{(k+1)} = \gamma^{(k)}\mathbf{p}^{(k)} - \mathbf{g}^{(k+1)}$, with $\mathbf{p}^{(0)} = -\mathbf{g}^{(0)}$, and coefficient $\gamma^{(k)}$ calculated by various means including the PRP, HS, and FR schemes [6].

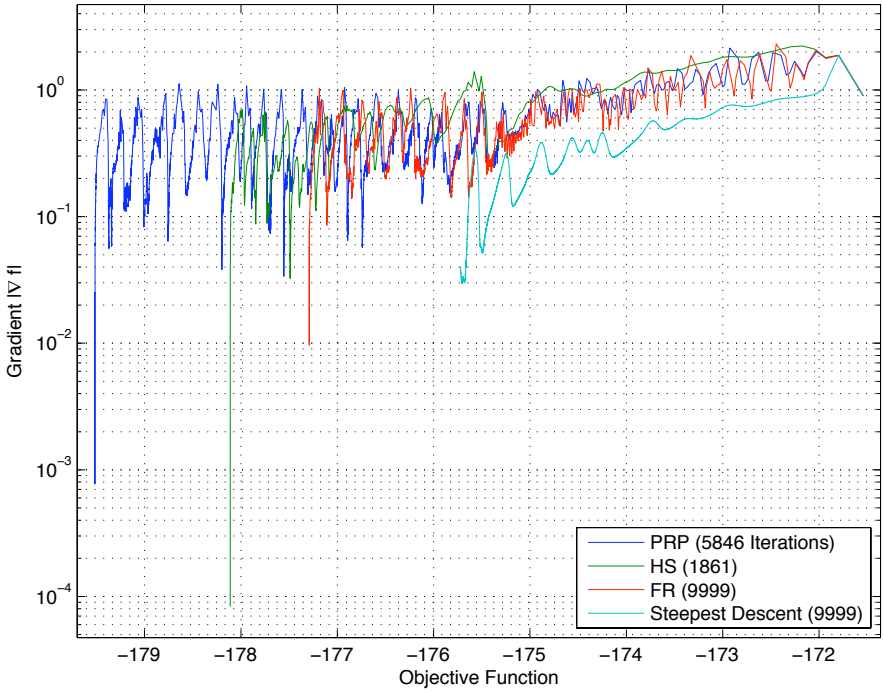

FIGURE 1: Conjugate gradient iteration trajectories on the objective function (1), for parameters $(\beta, \lambda) = (10, 0.0329)$, and $(\alpha, \eta) = (10^{-4}, 10^{-3})$.

3.2 Iteration performance and comparisons

Iteration trajectories for four different search direction update schemes are shown in the $(f, |\nabla f|)$ plane in Figure 1, with stopping tolerance of 10^{-12} on the relative absolute function difference $|\Delta f/f|$. The dominant feature of these trajectories is a sustained period of pronounced oscillation in the gradient norm, particularly for $\gamma^{(k)} \neq 0$, indicating a large proportion of iterations for which a gradient increase accompanied the objective function decrease. Essentially all of the objective function reduction work has been accomplished during this initial oscillatory phase of the trajectories, which offered little overall gradient reduction, with subsequent 'tails' of the trajectories effectively reducing the gradient alone. Such oscillatory gradient behaviour is not uncommon in unconstrained gradient descent optimisation, and was shown to be a sign of ill-conditioning for the steepest descent case ($\gamma^{(k)} = 0$) applied to a model convex quadratic function [8], which helps to explain the wide variations in trajectory behaviour evident in Figure 1.

The two best results, and the only ones to reach the stopping tolerance in less than 10000 iterations, PRP and HS, offer an interesting comparison. While the PRP scheme produced the lowest objective function, it did so at the expense of a higher final gradient and over three times the number of iterations for the HS scheme. The corresponding trajectory 'tails', where most of the gradient reduction was done, contain approximately 2000 and 500 iterations respectively for the PRP and HS results, the latter of which demonstrated a far swifter gradient decay rate per iteration. Although substantial gradient reductions were achieved, by factors of more than 1000, the final gradient norms seem slightly excessive to support declarations of local minima at this point. Nocedal et al. [8] reported similar cases, in which gradient norms were not reduced as much as desired, and suggested that the iteration path of an optimisation algorithm may determine the size of the final gradient norm. More generally, the gradient norm was cited as an unreliable measure of accuracy in these calculations.

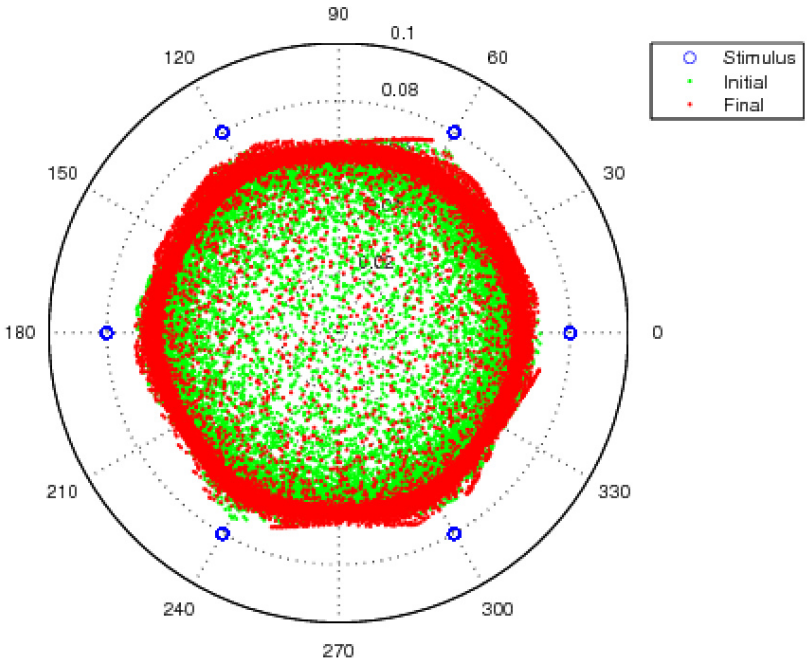

FIGURE 2: Polar plots of the initial and final orientation preference vectors, with corresponding stimulus points, for the PRP result from Figure 1, displaying a pronounced outward radial migration, representing enhanced coverage afforded by the objective function reduction.

TABLE 1: Relative percentage decreases in the objective function (1) and its two components, for each of the iteration trajectories shown in Figure 1.

	PRP	HS	FR	$\gamma^{(k)} = 0$
Continuity Decrease %	0.47	-0.10	-0.48	-1.67
Diversity Decrease	3.76	3.18	2.85	2.30
Objective Function Decrease	4.64	3.83	3.35	2.43

Relative reductions of the objective function, split into respective contributions from the two competing terms of (1), are given in Table 1 for the four schemes, in decreasing order. While each scheme reduced the diversity component, which dominated in every case, only the PRP result afforded a simultaneous reduction of the continuity component. For this result, Figure 2 gives polar plots of the initial and final orientation preference vectors [1], together with the associated stimulus points, showing enhanced coverage manifested as a pronounced radial migration of the calculated preferences towards their stimuli, creating a ring structure with clear hexagonal tendencies on its outer boundary. Complementary cortical plane views in Figure 3, displaying the related orientation preference maps [9], illustrate the accompanying continuity enhancement in dramatic fashion. Iso-orientation domains experienced a profound expansion to yield a stripe-like pattern with local plane wave character, presenting a stark contrast in smoothness between the initial and final states. Facilitating the formation of these large iso-orientation domains is a depletion of pinwheel structures through annihilation events, by which pairs of opposing pinwheels collide and merge to create local linear zones [9], reducing the pinwheel density to a third of its initial value in this case. Linear zones and pinwheel arrangements were observed by Bosking et al. [2], in a laboratory study exploring the horizontal connections established by neurons of known orientation preference.

The companion HS and FR results included for comparison in Figure 3 also underwent similar enlargement and smoothing processes, though clearly not to the extent seen in the PRP case, displaying a ranking according to their overall degree of objective function reduction. In all three cases, the difference between initial and final orientation preference states, as expressed in Frobenius norm, was larger than the initial value, producing relative changes of over 100% during the iteration span. This presents a sharp contrast to the corresponding objective function changes of below 5%.

The emergent stripe-like pattern of iso-orientation domains appearing in Figure 3 for the PRP result bears a striking resemblance to a particular result

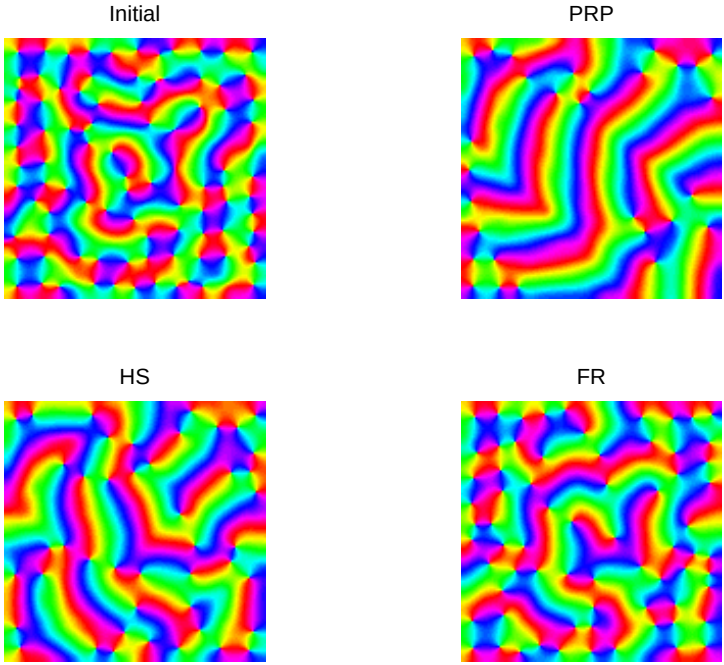

FIGURE 3: Final iterates from the conjugate gradient iteration trajectories in Figure 1, shown as π-periodic orientation preference maps in the cortical plane, together with the common initial condition. The `hsv` colourmap cycles from red ($-90°$), through yellow, green, cyan, blue, magenta and back to red.

by Wolf and Geisel [9], generated via a dynamical system approach. Considering the coordinated development of orientation and ocular dominance columns, with a focus on the rearrangement of iso-orientation domains under varying degrees of ocular dominance segregation, ranging from none to strong, the study [9] demonstrated an ability of ocular dominance to suppress pinwheel annihilation. In the absence of ocular dominance columns, as for the calculations of Figure 1, copious pinwheel annihilation was observed, in agreement with the PRP result of Figure 3. Such agreement also supports statements by Wolf and Geisel [9] regarding an independence of model parameters and details, for the observed ocular dominance effects. Under similar conditions, excluding ocular dominance, an unspecified "efficient minimization method based on Cholesky factorization", applied without numerical supporting evidence by Carreira–Perpiñán, Lister and Goodhill [3] produced a very different picture with considerably higher pinwheel density, presenting a stark contrast with the PRP result in Figure 3, and with Wolf and Geisel's result [9]. This substantial discrepancy in orientation preference map structure raises immediate questions concerning the numerical method used by Carreira–Perpiñán, Lister and Goodhill [3], and whether or not their calculated result represents a converged local minimum of the objective function.

3.3 Gradient spikes

Distinct spiking activity observed in the iteration trajectories of Figure 1 must be playing a certain role in the spatial evolution process. To examine what this may be signalling, some early iteration behaviour has been captured in Figure 4, comprising the gradient profile for the first 250 iterations accompanied by three snapshots of the orientation preference map, after 100, 125 and 175 iterations. Two prominent spikes occurred during this interval, the first of which peaks just below 125 iterations, bringing attention to the first two orientation preference maps in the figure. Comparing these reveals a pinwheel annihilation event taking place in the lower left quadrant, whereby two opposing pinwheels visible at iteration 100 collided to yield a

locally linear iso-orientation zone at iteration 125, as also seen by Wolf and Geisel [9]. The following spike, peaking just before iteration 175, coincides with an event involving a quartet of pinwheels residing in the upper left quadrant, comprising two pairs of opposing pinwheels. Individual pinwheels in each pair are connected by a periodic interface line, made visible by the `jet` colourmap as a blue-red transition, across which the preferred orientation jumps by 180°. Close inspection of the results at iterations 125 and 175 reveals that the interconnections have actually exchanged, leaving the four pinwheels intact and breaking a region of high positive values (red) to allow the union of two neighbouring regions of large negative (blue) angles.

At least two different evolution processes appear to be leaving strong signatures in the iteration trajectories, revealing another facet of the oscillatory gradient behaviour which has also been linked to ill-conditioning [8]. After the oscillation has subsided, in the trajectory tails of Figure 1, this suggests that little further structural evolution should occur. For the PRP result, the relative change between iteration 4000 and the final result is just under 13%.

4 Summary and conclusions

For a four dimensional receptive field comprising orientation preference and retinal coordinates, conjugate gradient iteration has succeeded in transforming a pinwheel-rich initial state to yield a very similar orientation preference map structure to that of a previous study utilising a dynamical system on a coarser cortical mesh at different parameter settings, reinforcing some of its key conclusions. To achieve this agreement required a sufficiently low objective function value, only attained after experimenting with various search direction updates and line search parameter settings. Persistent gradient oscillation in the iteration trajectories was linked to ill-conditioning and certain spatial evolution processes involving pinwheel structures in the orientation preference map.

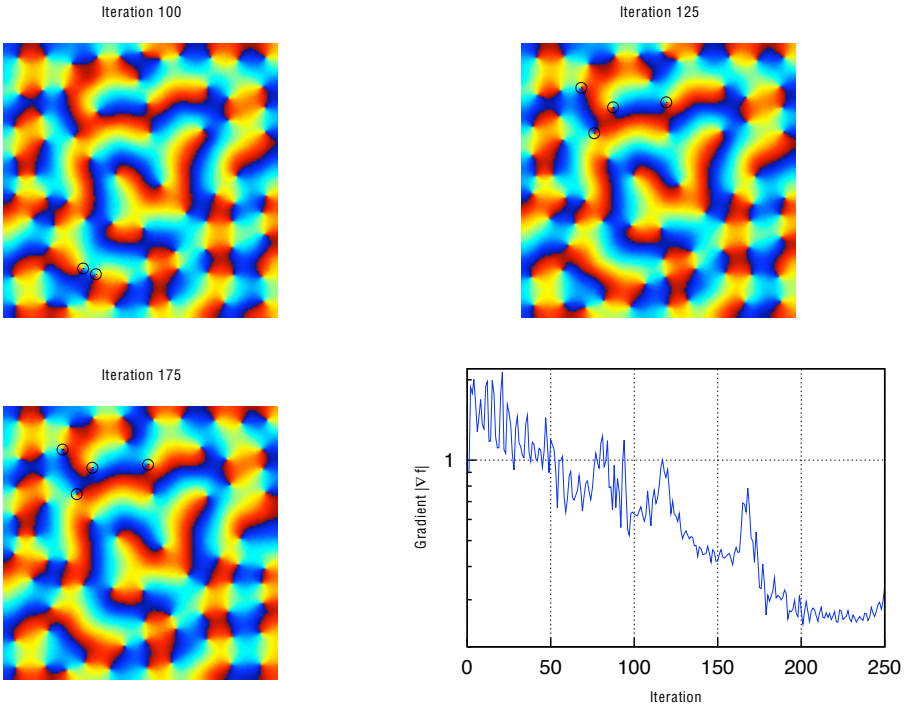

FIGURE 4: Computed orientation preference maps after 100, 125 and 175 PRP iterations, showing a pinwheel annihilation event in the lower left quadrant between iterations 100 and 125, circled, followed by a different event involving a quartet of pinwheels in the upper left quadrant, also circled. Both events appear to have left strong spike signatures in the gradient profile.

Before extending the calculations to higher dimensional receptive fields, beginning with the inclusion of ocular dominance as the fifth dimension, additional four dimensional studies will be carried out via quasi Newton methods with reduced Hessian approximations [7]. At little extra storage cost, these methods offer potential performance advantages over the conjugate gradient iteration, and their iteration trajectories should provide interesting comparisons, particularly with regard to the degree of gradient reduction.

Acknowledgements The author is grateful to G. Goodhill at QBI for supplying an initial condition, stimulus points, and program fragments.

References

[1] G. Blasdel, Orientation Selectivity, Preference, and Continuity in Monkey Striate Cortex, *The J. Neuroscience* **12**, 8 p3139–3161 (1992). http://www.jneurosci.org/cgi/content/abstract/12/8/3139 C78, C83

[2] W. H. Bosking et al., Orientation Selectivity and the Arrangement of Horizontal Connections in Tree Shrew Striate Cortex, *The Journal of Neuroscience* **17**, 6 p2112–2127 (1997). http://www.jneurosci.org/cgi/content/abstract/17/6/2112 C76, C83

[3] M.Á. Carreira–Perpiñán, R. J. Lister and G. J. Goodhill, A Computational Model for the Development of Multiple Maps in Primary Visual Cortex, *Cerebral Cortex* **15**, 8 p1222–1233 (2005). doi:10.1093/cercor/bhi004 C76, C85

[4] R. Durbin and G. Mitchison, A dimension reduction framework for understanding cortical maps, *Nature* **343**, p644–647 (1990). doi:10.1038/343644a0 C76, C78

[5] R. Fletcher, *Practical Methods of Optimization*, Wiley (1987). C79

[6] J. Gilbert and J. Nocedal, Global Convergence Properties of Conjugate Gradient Methods for Optimization, *SIAM J. Opt.* **2**, 1 p 21–42 (1992). doi:10.1137/0802003 C79

[7] P. E. Gill and M. W. Leonard, Limited-Memory Reduced-Hessian Methods for Large-Scale Unconstrained Optimization, *SIAM J. Optimization* **14**, 2 p 380–401 (2003). doi:10.1137/S1052623497319973 C88

[8] J. Nocedal et al., On the Behavior of the Gradient Norm in the Steepest Descent Method, *Comp. Optimization and Appl.* **22**, 1 p5–35 (2002). doi:10.1023/A:1014897230089 C77, C81, C86

[9] F. Wolf and T. Geisel, Spontaneous pinwheel annihilation during visual development, *Nature* **395**, 3 p73–78 (1998). doi:10.1038/25736 C76, C77, C78, C83, C85, C86

[10] N. V. Swindale and H. Bauer, Application of Kohonen's self-organizing feature map algorithm to cortical maps of orientation and direction preference, *Proc. R. Soc. Lond. B* **265**, p827–838 (1998). doi:10.1098/rspb.1998.0367 C76, C78

[11] *Numerical Recipes Online*, http://www.nr.com. C79

Author address

1. **D. H. Smith**,
mailto:dherschelsm@yahoo.com