

A multigrid approach to visual cortex mapping

David H. Smith¹

(Received 30 January 2011; revised 11 June 2011)

Abstract

Standard unconstrained optimisation methods, utilising conjugate gradient and limited memory reduced Hessian search direction updates, are implemented in a multigrid mode on an objective function minimisation model of visual cortex mapping. The multigrid operations, executed on a hierarchy of four grids, easily outpace fine grid iterations, and penetrate deeper into the objective function.

Contents

1	Introduction	C304
2	Numerical methods	C305
2.1	The objective function	C305
2.2	Iteration details	C306
2.3	Multigrid implementation	C307

<http://anziamj.austms.org.au/ojs/index.php/ANZIAMJ/article/view/3953>
gives this article, © Austral. Mathematical Soc. 2011. Published July 1, 2011; amended July 3, 2011. ISSN 1446-8735. (Print two pages per sheet of paper.) Copies of this article must not be made otherwise available on the internet; instead link directly to this URL for this article.

<i>1</i>	<i>Introduction</i>	C304
3	Numerical results	C308
3.1	Fine grid calculations	C310
3.2	Acceleration by multigrid	C310
4	Summary and conclusions	C315
	References	C316

1 Introduction

Multigrid methods offer substantial performance enhancements in a wide variety of computational settings, from elliptic solvers in atmospheric flows [8] to shape optimisation of turbine blades [5]. Visual cortex mapping calculations, implemented under the elastic network formulation [4, 6], are also well placed to reap the benefits of multigrid. This particular model involves the unconstrained minimisation of a certain objective function, the arguments of which comprise neuronal preferences for various visual attributes including orientation and position, defined on a regular mesh. Nonlinear conjugate gradient iteration has successfully reduced this objective function to levels which provide results agreeing qualitatively with those produced by an equivalent dynamical systems approach [12, 13], at the cost of rather slow convergence. The main objective of this study is to revisit my previous calculations [12], armed with an alternative quasi-Newton search direction update [3] and multigrid, in pursuit of enhanced performance.

The fine grid under consideration is a 128×128 cortex mesh, holding four variables at each grid point, making a total of 2^{16} . This becomes the top of a four level grid hierarchy, with successively coarser levels simply halving the number of points in each direction. Data transfer between grids plays a key role, fulfilled by four different prolongation operators for coarse to fine transfer, and direct copying for the reverse direction. After reviewing the basic iteration schemes and demonstrating their fine grid behaviour, extensions to

multigrid are outlined. A set of 64 numerical experiments will be presented, comparing different aspects via a specifically chosen compact performance measure in a single plane view.

2 Numerical methods

The calculations described here fall into the category of unconstrained minimisation, applied to a nonlinear objective function of 2^{16} variables. Memory limitations prohibit the storage of full Hessian matrices of this size, necessitating economy storage options. The term “visual cortex map” refers to the pattern of calculated preferences, such as orientation angle, spatially evolving across the cortex lattice.

2.1 The objective function

The visual cortex is treated as a two dimensional sheet onto which response properties are mapped, according to the elastic network formulation [6]. For a discrete cortex lattice, with receptive field vector \mathbf{u}_j defined at each lattice location j , the objective function

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

where \mathbf{u} is the concatenation of all \mathbf{u}_j , and \mathbf{v}_i is a corresponding visual stimulus point, similarly concatenated in \mathbf{v} , represents conflicting biological demands of continuity and coverage on the visual cortex [4, 6]. Quoting from Carreira–Perpiñán et al. [4], “This is a feature space model which works by minimizing an objective function that explicitly trades off coverage versus continuity constraints”. Continuity is enforced by the sparse matrix \mathbf{A} acting on the four nearest neighbours of each cortical point to give a sum of squared differences operation. Coverage is represented in the second term of (1) via

squared differences between receptive fields and associated visual stimuli at every cortex point. Parameters λ and β denote the receptive field size and lateral interaction strength respectively, as detailed by Wolf and Geisel [13].

The four dimensional receptive field under consideration comprises two retinal position coordinates and two orientation coordinates,

$$\mathbf{u}_j = [x_j \quad y_j \quad r_j \cos 2\theta_j \quad r_j \sin 2\theta_j]^\top,$$

in which (x_j, y_j) are the centre of the receptive field in visual space, θ_j is the preferred angle of orientation, and r_j is the associated selectivity [6]. Stimulus points \mathbf{v}_i , comprising a set of size 2400, and parameter values for (λ, β) , are all taken from my earlier work [12].

2.2 Iteration details

At iteration k , a line search [10] attempts to minimise the objective function along the current search direction \mathbf{p}_{k-1} ,

$$\phi(s) = f(\mathbf{u}^{(k-1)} + s\mathbf{p}_{k-1}),$$

producing the next iterate

$$\mathbf{u}^{(k)} = \mathbf{u}^{(k-1)} + s_{k-1}\mathbf{p}_{k-1}, \quad k = 1, 2, \dots$$

The line search algorithm is based on quadratic and cubic interpolants of the objective function along the search direction [10], and terminates when the average and instantaneous rates of change meet certain tolerances according to the Goldstein–Wolfe criteria [7, 9],

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

with parameters $0 < \alpha < \eta < 1$.

Two different search direction updates are employed: the first being the best performer from previous conjugate gradient calculations [12], combining the

new and previous gradients $\mathbf{g}_k = \nabla f(\mathbf{u}^{(k)})$, \mathbf{g}_{k-1} , and the previous search direction \mathbf{p}_{k-1} via the Polak–Ribière–Polyak (PRP) update [9],

$$\mathbf{p}_k = \frac{(\mathbf{g}_k - \mathbf{g}_{k-1})^\top \mathbf{g}_k}{\mathbf{g}_{k-1}^\top \mathbf{g}_{k-1}} \mathbf{p}_{k-1} - \mathbf{g}_k. \quad (3)$$

The second is a limited memory version of the BFGS inverse Hessian update (L-BFGS), based on a set of secant pairs, $\mathbf{y}_k = \mathbf{g}_{k+1} - \mathbf{g}_k$ and $\mathbf{z}_k = \mathbf{u}^{(k+1)} - \mathbf{u}^{(k)}$, generated from recent iterations [3, 7]

$$\mathbf{p}_k = -\gamma_k \mathbf{g}_k - \mathbf{Z}_k \left[\mathbf{R}_k^{-\top} (\mathbf{D}_k + \gamma_k \mathbf{Y}_k^\top \mathbf{Y}_k) \mathbf{R}_k^{-1} \mathbf{Z}_k^\top - \gamma_k \mathbf{R}_k^{-\top} \mathbf{Y}_k^\top \right] \mathbf{g}_k + \gamma_k \mathbf{Y}_k \mathbf{R}_k^{-1} \mathbf{Z}_k^\top \mathbf{g}_k. \quad (4)$$

Formula (4) is a low budget approximation to the acclaimed Newton search direction [7], avoiding storage of the full Hessian matrix, which is prohibitively large in these calculations. Columns of the matrices \mathbf{Y}_k and \mathbf{Z}_k contain a set of recent secant pairs, while curvature information is furnished by the upper triangular matrix \mathbf{R}_k ,

$$\mathbf{R}_k(i, j) = \begin{cases} \mathbf{z}_{i-1}^\top \mathbf{y}_{j-1}, & i \leq j, \\ 0, & i > j, \end{cases}$$

with \mathbf{D}_k holding its diagonal elements, and the scalar $\gamma_k = \mathbf{y}_k^\top \mathbf{z}_k / \mathbf{y}_k^\top \mathbf{y}_k$. At the first iteration, $k = 0$, when no previous search direction is available, the steepest descent direction is applied, along the negative gradient.

2.3 Multigrid implementation

As the quantities of interest in these calculations reside on a regular mesh, or matrix, they naturally lend themselves to the multigrid methodology, involving a judicious combination of operations on a hierarchy of meshes, with appropriate transfer between levels playing a key role [2]. With little extra effort, the conjugate gradient and quasi-Newton iteration schemes are implemented in a multigrid setting, using the same set of stimulus points.

Consider a hierarchy of four grids, the top level comprising the original 128×128 cortical mesh, while subsequent coarser grids simply halve the number of points in each direction, giving 64×64 , 32×32 , and finally 16×16 on the bottom level. A multigrid V cycle commences by executing a small number of iterations on the top grid, the results of which are restricted to the next coarser grid, either by direct copying or a weighted average process [2]. Invoking recursion, the same routine is then called, with more iterations being done as the grid becomes coarser, before finally prolongating the result back to the fine level and iterating again to complete the cycle. The prolongation is achieved by an interpolation operator, from simple linear to various higher order polynomials and splines [2]. Recursive W cycles involve a second restriction to the coarse level, generating additional effort on the coarser grids, with potential performance improvements.

There is considerable scope for tuning in multigrid calculations, limited in this study to prolongation operators, search direction updates, and exit tolerance applied to the relative absolute difference in the objective function. Fixed iteration limits of 16, 32, 64 and 128 were imposed, in order from fine to coarse, so that each V cycle involves a maximum of 32 fine grid iterations, or 48 for a W cycle. Direct copying is applied for restriction, while prolongation is implemented with a piecewise cubic Hermite (PCH) spline, cubic spline, local cubic interpolation, or local linear interpolation, as supplied by the `Octave` numerical computation package [11]. Line search exit parameters for (2) are fixed at $(\alpha, \eta) = (0.0001, 0.001)$, and all available secant pairs are used in (4), which makes a comfortable 16 on the fine grid. Search direction or gradient information is not transferred between levels, giving ‘kick starts’ on each level with steepest descent directions.

3 Numerical results

I report on the fine grid calculations, and the performance improvements achieved by using a multigrid strategy.

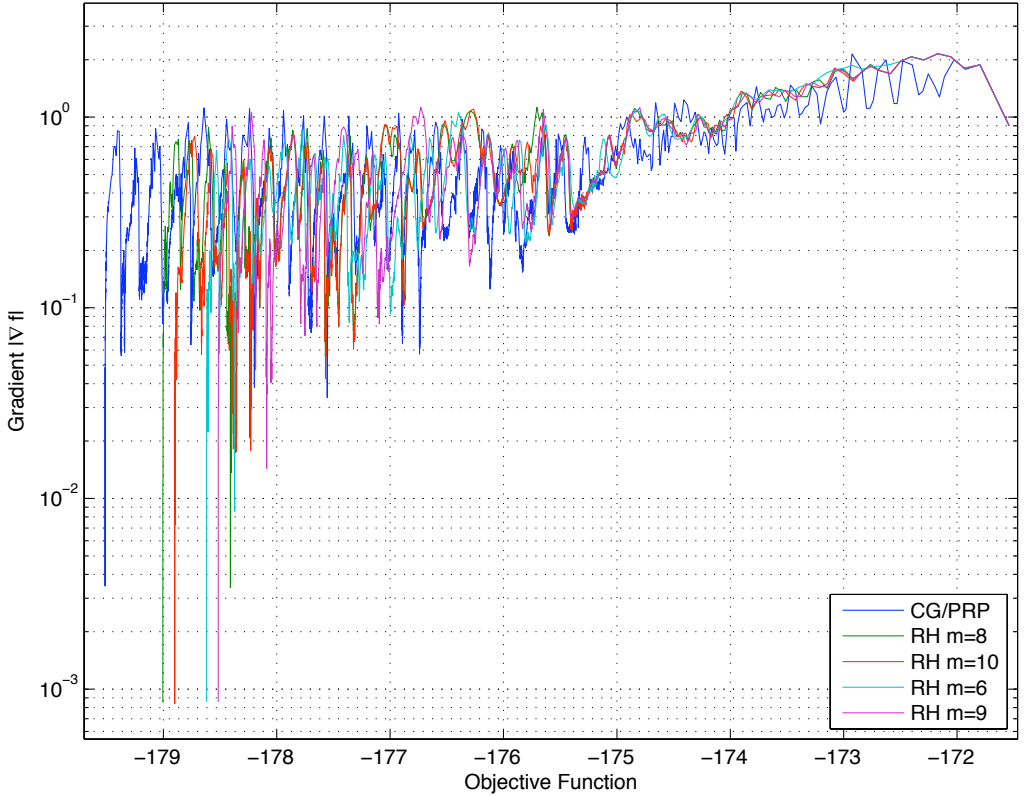


FIGURE 1: Comparison of previous best CG/PRP result from [12] with nearest rivals generated by quasi-Newton reduced Hessian (RH), with m denoting the number of secant pairs in the Hessian matrix reconstruction. The common initial condition is that I previously used [12].

3.1 Fine grid calculations

Previous calculations on a 128×128 cortex mesh, employing conjugate gradient iteration with a variety of search direction update schemes [12], were characterised by long and tedious trajectories, taking thousands of iterations to converge. Best results in [12], achieved with the PRP search direction update (3), are compared with some L-BFGS results in the iteration trajectories of Figure 1, showing continued superiority of the CG/PRP scheme. Also to note in this diagram is the oscillatory gradient behaviour, as previously observed [12], and local minima trapping.

3.2 Acceleration by multigrid

Fine grids are often allied with slow convergence [2], as witnessed in the trajectories of figure 1, requiring thousands of iterations to converge on the original 128×128 grid. As an initial demonstration of the multigrid implementation, consider the previous best CG/PRP fine grid result from Figure 1, which has been executed in V cycle mode with four different prolongation operators, producing the results displayed in Figure 2. For 50 V cycles, this diagram gives objective function profiles on the top level grid, showing deeper penetration of the multigrid cycles in comparison to the original fine grid profile also included. This is dramatically demonstrated in the first cycle, where the multigrid curves all break away after the common initial 16 iterations, revealing the valuable contribution from the coarser grids.

In terms of overall objective function reduction, the multigrid operations achieve in roughly 10 cycles what the fine grid iterations achieve in over 100 equivalent cycles, at 32 iterations per cycle, representing substantial savings. However there appear to be some negative side effects, most obviously a ‘stalling’ phenomenon visible in later cycles, which return larger objective function values than those at the cycle start. This is particularly pronounced for the local linear prolongation, appearing as early as the fifth

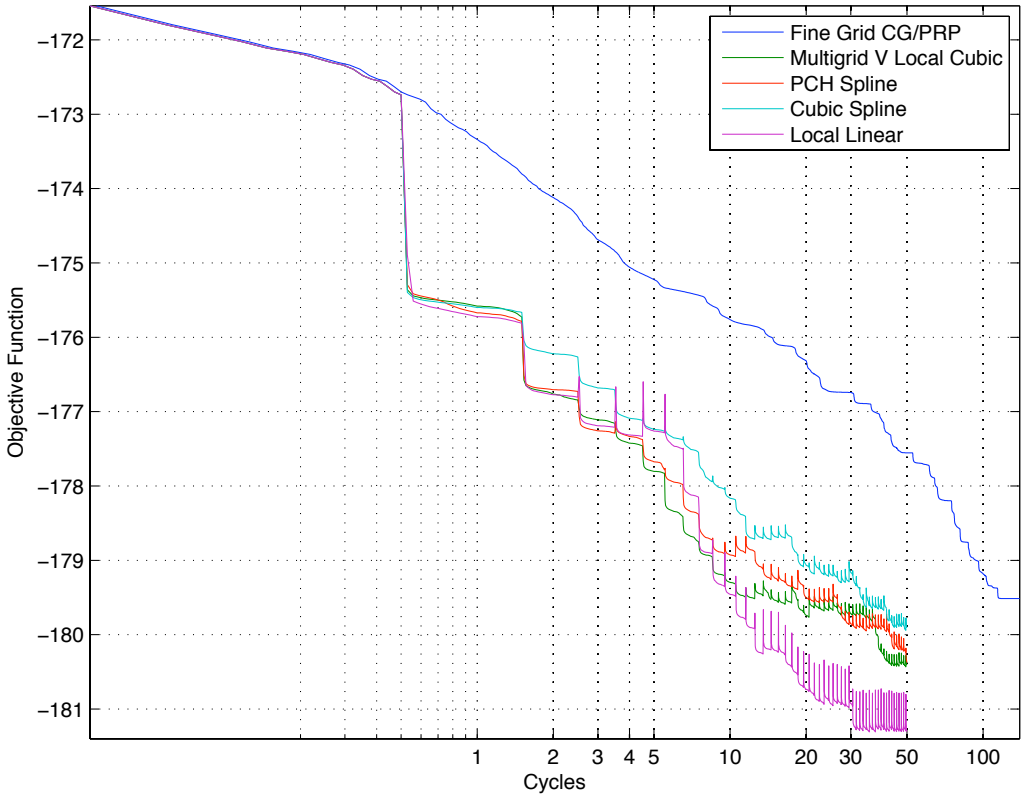


FIGURE 2: Multigrid V cycle implementation of the CG/PRP scheme, using four different prolongation operators at exit tolerance 10^{-10} , with the original fine grid result, horizontally scaled to an equivalent number of cycles.

cycle, accompanied by large spikes upon return to the top grid. Ironically, it is also the local linear prolongator that returns the best overall result after 50 cycles, penetrating well below its counterparts amid some intense spiking activity. On closer inspection, this result actually holds the lead after 10 cycles, subsequently reaching its lowest depth near 30 cycles, with negligible overall progress in the final 20. Spiking is also present in the objective function gradient, somewhat familiar in the light of Figure 1 and previous studies [12], to a degree that precludes any appreciable gradient reduction during the 50 cycles. Not surprisingly then, convergence to the specified tolerance has not occurred; however, what is more important at this stage is the deeper objective function penetration, especially its implication for the resulting orientation preference map structure.

Pinwheels, or orientation vortices [6], are singularities at which no preference occurs, mathematically defined as the intersection of zero contours for the cartesian orientation preference components [13]. These key structures have become a regular focus of experimental and computational studies [1, 12, 13], once again attracting attention in this study. A steady pinwheel depletion was observed previously [12] as the iterations progressed, ultimately yielding large stripe-like iso-orientation domains with local plane wave character, in qualitative agreement with the dynamical systems calculations of [13]. Further objective function reductions beyond those achieved in [12], afforded by the multigrid results of Figure 2, indeed brought fewer attendant pinwheels, in line with the expected depletion behaviour. This underlying connection between objective function and pinwheel density points naturally to a certain visualisation device and performance indicator based on these two quantities, to expedite comparison of different multigrid results.

The four multigrid results appearing in Figure 2 belong to a set of 64 numerical experiments involving both search direction updates (3) and (4), V and W cycles, four prolongation operators, and four exit tolerances, between 10^{-9} and 10^{-12} . Using the final objective function value and number of pinwheels present in the associated orientation preference map, Figure 3 presents a graphical summary of the entire set, also including some of the

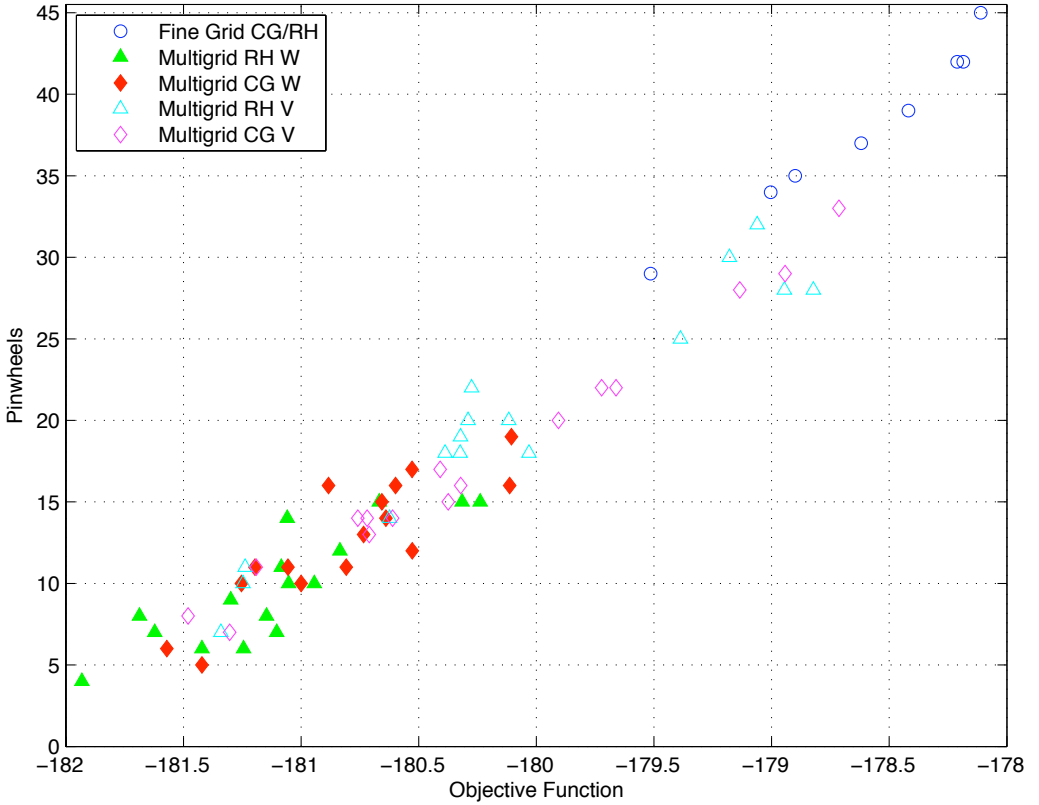


FIGURE 3: Summary of 64 multigrid experiments, expressed as final objective function value and corresponding number of pinwheels in the orientation preference map, accompanied by some fine grid results. The experiments comprise 50 multigrid cycles with V and W configurations, four exit tolerances, four prolongation operators, and two search direction updates.

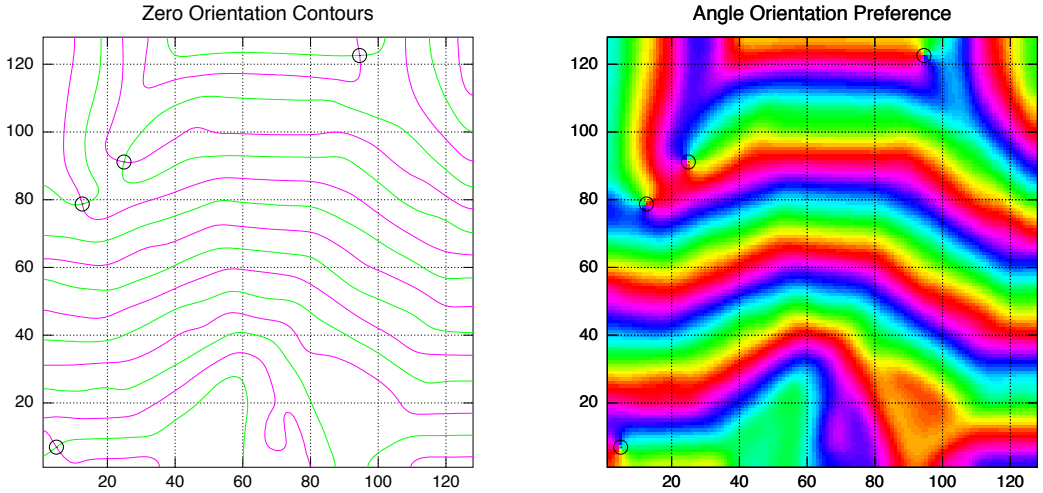


FIGURE 4: Calculated orientation preference map for the overall best result from Figure 3, produced by multigrid W cycles and reduced Hessian search direction updates. The left panel gives zero contours of the cartesian orientation preference components, with pinwheels at the intersection points, while the right panel gives pinwheels with preference angles (π periodic) using the HSV colormap (red -90° , yellow, green, cyan, blue, magenta).

leading fine grid results for comparison. A striking linear tendency is apparent in this view, with superiority of W cycles evident by the dominance of these points at the lower left of the diagram. The V cycle results have dispersed further along the linear corridor, being outperformed by fine grid iteration on a small number of occasions. In terms of search direction updates, W cycles favoured reduced Hessian while V cycles favoured conjugate gradient. Regrouping the results in Figure 3 by prolongation operators sees PCH ahead on average, and similarly for exit tolerance, the value 10^{-10} held a narrow lead.

The overall best performance in Figure 3, possessing only four pinwheels in its orientation preference map, was claimed by multigrid W cycles operating

with reduced Hessian search direction update, local cubic prolongation, and exit tolerance of 10^{-11} . Two alternative versions of the resulting orientation preference map are displayed in Figure 4, one with zero contours and the other with angles, demonstrating the characteristic plane wave structure seen previously [12, 13], with even larger iso-orientation domains. Like the multigrid results in Figure 2, this result has not yet converged on the fine grid; however, additional fine grid iteration essentially reduces the gradient only, achieving convergence along vertical ‘tails’ as in the trajectories of Figure 1. Slight migration of the four pinwheels also took place during the iteration, with no further depletion occurring, showing a relative insensitivity of pinwheel density to gradient. Such a low pinwheel density in Figure 4 leads to the conjecture that, with further targeted numerical experimentation, all pinwheels can be removed to deliver a plane wave structure, placing in question the model’s ability to support pinwheel structures in this particular four dimensional configuration, at the associated parameter settings.

4 Summary and conclusions

By treating a 128×128 cortex mesh as the top of a four level hierarchy, multigrid implementation of standard unconstrained minimisation schemes on a visual cortex model delivered deeper objective function penetration than that of exclusive fine grid operations, at a fraction of the computational cost. Pinwheel densities in the resulting orientation preference maps were reduced to an extent that challenges the model’s capacity to support these structures, in its current configuration.

Scope for improvement exists, particularly with respect to the gradient behaviour, which underwent oscillation and little overall reduction during the multigrid cycles. Transferring gradient and search direction information between grids, in addition to the working variables, is an obvious option to consider in this regard. An intermittent stalling phenomenon, which reverses progress in some cycles, points to the prolongation operators for coarse to

fine transfers. Restriction operators could also be enhanced by employing weighting of neighbours rather than direct copying.

Having established a working multigrid system on four dimensional receptive fields, future studies will enhance and apply this methodology on a more computationally demanding five dimensional case, which incorporates ocular dominance with the existing orientation and position attributes. An expected suppression of the pinwheel depletion [13] will be explored along two initial avenues, one starting from the same pinwheel-rich initial state of the current study, and the other starting from a heavily depleted state produced by the multigrid schemes, the latter of which should see a return of pinwheels as the objective function is reduced.

References

- [1] 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>
C312
- [2] W. L. Briggs, *A Multigrid Tutorial*, SIAM, Philadelphia (1988). C307, C308, C310
- [3] R. H. Byrd et al., Representations of quasi-Newton matrices and their use in limited memory methods, *Mathematical Programming* **63**, 4 p129–156 (1994). doi:10.1007/BF01582063 C304, C307
- [4] 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 C304, C305

- [5] T. Dreyer et al., Multigrid optimization in applications, *J. Comp. Appl. Math.* **120** p67–84 (2000). doi:10.1016/S0377-0427(00)00304-6 C304
- [6] R. Durbin and G. Mitchison, A dimension reduction framework for understanding cortical maps, *Nature* **343**, p644–647 (1990). doi:10.1038/343644a0 C304, C305, C306, C312
- [7] R. Fletcher, *Practical Methods of Optimization*, Wiley (1987). C306, C307
- [8] S. Fulton et al., Multigrid methods for elliptic problems: A review, *Monthly Weather Review* **114** p943–959 (1986). [http://dx.doi.org/10.1175/1520-0493\(1986\)114<0943:MMFEPA>2.0.CO;2](http://dx.doi.org/10.1175/1520-0493(1986)114<0943:MMFEPA>2.0.CO;2) C304
- [9] J. Gilbert and J. Nocedal, Global convergence properties of conjugate gradient methods for optimization, *SIAM J. Opt.* **2**, 1 p21–42 (1992). doi:10.1137/0802003 C306, C307
- [10] *Numerical recipes online*, <http://www.nr.com>. C306
- [11] *Octave numerical computation software* <http://www.octave.org>. C308
- [12] D. H. Smith, Visual cortex mapping by conjugate gradients, *ANZIAM J.* **50** (CTAC2008) pC75–C89, (2008). <http://anziamj.austms.org.au/ojs/index.php/ANZIAMJ/article/view/1373> C304, C306, C309, C310, C312, C315
- [13] F. Wolf and T. Geisel, Spontaneous pinwheel annihilation during visual development, *Nature* **395**, 3 p73–78 (1998). doi:10.1038/25736 C304, C306, C312, C315, C316

Author address

1. David H. Smith,
<mailto:dherschelsm@yahoo.com>