

Fast adaptive estimation of multidimensional psychometric functions

Christopher DiMattina

Computational Perception Laboratory,
Department of Psychology, Florida Gulf Coast University,
Fort Myers, FL, USA



Recently in vision science there has been great interest in understanding the perceptual representations of complex multidimensional stimuli. Therefore, it is becoming very important to develop methods for performing psychophysical experiments with multidimensional stimuli and efficiently estimating psychometric models that have multiple free parameters. In this methodological study, I analyze three efficient implementations of the popular Ψ method for adaptive data collection, two of which are novel approaches to psychophysical experiments. Although the standard implementation of the Ψ procedure is intractable in higher dimensions, I demonstrate that my implementations generalize well to complex psychometric models defined in multidimensional stimulus spaces and can be implemented very efficiently on standard laboratory computers. I show that my implementations may be of particular use for experiments studying how subjects combine multiple cues to estimate sensory quantities. I discuss strategies for speeding up experiments and suggest directions for future research in this rapidly growing area at the intersection of cognitive science, neuroscience, and machine learning.

Introduction

Several recent studies in vision science have examined how observers combine information from multiple stimulus dimensions to make perceptual decisions (Landy & Kojima, 2001; Hillis, Watt, Landy, & Banks, 2004; Knill & Pouget, 2004; Trommershauser, Kording, & Landy, 2011). Combination of multiple sensory cues is essential for natural perception, as demonstrated in Figure 1, where we see that detection and localization of this occlusion edge requires combining multiple sources of information like luminance, color, and texture. Doing psychophysics with multidimensional stimuli poses a significant methodological

challenge, and therefore traditionally most psychophysical studies have considered only variations along a single stimulus dimension, holding other stimulus features fixed (Kingdom & Prins, 2010; Lu & Doshier, 2013). While this traditional approach has proven highly effective for simple, artificial stimuli like bars and gratings defined by only a few parameters (Skottun, Bradley, Sclar, Ohzawa, & Freeman, 1987; Vogels & Orban, 1990), it may not be as effective for complex naturalistic stimuli defined by multiple interacting feature dimensions (Landy & Kojima, 2001; McGraw, Whitaker, Badcock, & Skillen, 2003; Ing, Wilson, & Geisler, 2010; DiMattina, Fox, & Lewicki, 2012; Zavitz & Baker, 2014), particularly in cases where the combination of stimulus features is nonlinear.

Quantifying perception of multidimensional stimuli often gives rise to models that have multiple free parameters which must be estimated from experimental data. In this paper, I will use the term *multidimensional* to refer to the case where there are multiple stimulus dimensions which give rise to models that have a relatively large number of free parameters. The major problem posed by the need to identify multidimensional sensory-processing models from experimental data is the ability to collect sufficient data for attaining accurate estimates of model parameters in the time available to work with a subject. One approach to speeding up the collection of sensory data is to design stimuli adaptively so that at each trial, one presents stimuli optimized for the goal of accurate parameter estimation (Chaloner & Verdinelli, 1995; Lewi, Butera, & Paninski, 2009; DiMattina & Zhang, 2011). This general approach of adaptively choosing stimuli during an experiment to maximize some utility function goes by a variety of names, including *adaptive design optimization* (Cavagnaro, Myung, Pitt, & Kujala, 2010) and *optimal experimental design* (Atkinson, Donev, & Tobias, 2007; DiMattina & Zhang, 2011), and many different algorithms of this kind have

Citation: DiMattina, C. (2015). Fast adaptive estimation of multidimensional psychometric functions. *Journal of Vision*, 15(9):5, 1–20, doi:10.1167/15.9.5.

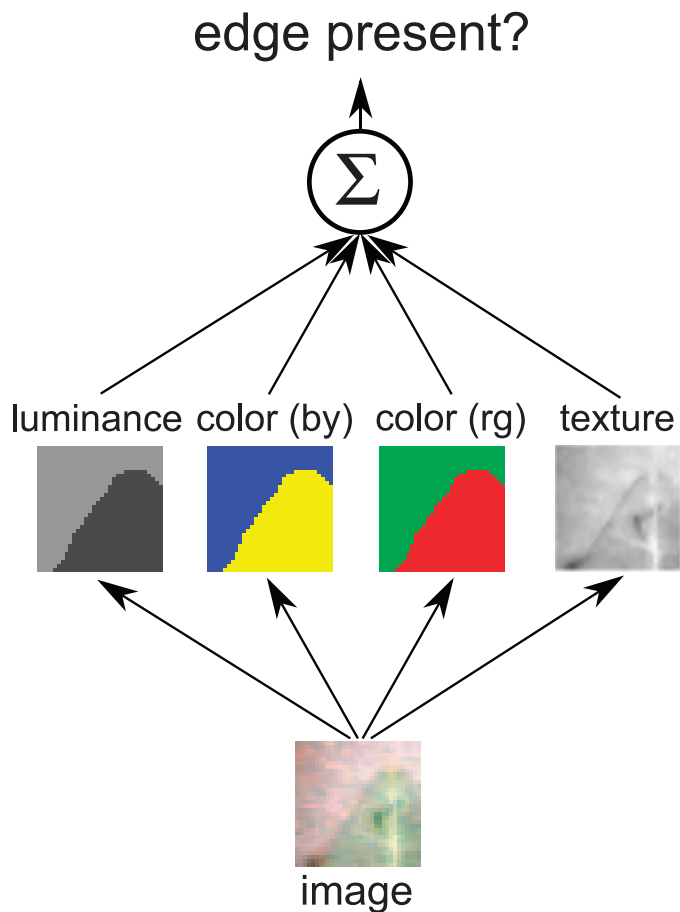


Figure 1. Complex natural stimuli like this occlusion edge are defined by multiple cues which must be integrated to make a perceptual decision.

been proposed and applied in experiments to accelerate the process of estimating psychometric function parameters (Hall, 1981; Watson & Pelli, 1983; Kontsevich & Tyler, 1999; Kujala & Lukka, 2006; Lesmes, Lu, Baek, & Albright, 2010; Prins, 2013b). However, with few exceptions (Kujala & Lukka, 2006; Lesmes et al., 2010), nearly all of these procedures have been applied to estimating psychometric functions defined in 1-D stimulus spaces.

In this computational-methods study, I analyze three efficient implementations of the well-studied Ψ procedure (Kontsevich & Tyler, 1999) which generalize well to identifying multidimensional psychometric models. One of these implementations (Prior- Ψ) has been applied in previous work (Kujala & Lukka, 2006), while two of them (Lookup- Ψ and Laplace- Ψ) represent (to the best of my knowledge) novel proposals for psychophysical experiments. I demonstrate in simulated psychophysical experiments that these implementations offer a substantial speedup over the original grid-based implementation (Grid- Ψ), and make it possible to quickly estimate the parameters of multidimensional psychometric models using standard

laboratory computers and software packages. I demonstrate how my methods may be of particular use for studies of how subjects integrate multiple cues when making perceptual decisions. Finally, I point to directions for future research and provide code online for implementing these procedures.

Methods and results

Estimating psychometric functions

The psychometric function

In psychophysical experiments, the main object of study is to characterize the parameters θ of some psychometric function $F(\mathbf{x}, \theta)$ mapping stimulus parameter values $\mathbf{x}$ to the range $[0, 1]$. The goal of characterizing this function's parameters θ is to accurately model the process generating the data. For an n -alternative forced-choice experiment, the probability Ψ of a correct response given stimulus parameters $\mathbf{x}$ is given by

$$\Psi(\mathbf{x}, \theta, \pi_c, \pi_l) = (1 - \pi_l) \left(\pi_c + (1 - \pi_c) F(\mathbf{x}, \theta) \right) + \pi_l \pi_c, \quad (1)$$

where π_l is the lapse rate and π_c is the chance of obtaining a correct response by guessing (Kuss, Jäkel, & Wichmann, 2005). One popular choice of the psychometric function is given by

$$F(x, \theta) = \sigma(\theta_0 + \theta_1 x), \quad (2)$$

where

$$\sigma(u) = \frac{1}{1 + e^{-u}} \quad (3)$$

is the logistic or sigmoidal function commonly used in machine learning and neural modeling (Murphy, 2012). Although there are many choices of sigmoid-shaped curves that are commonly used in modeling psychophysical data, all of them yield very similar estimates of important psychophysical parameters like sensory thresholds (Kingdom & Prins, 2010). A different parameterization of the psychometric function in Equation 2 that makes more explicit the psychophysical parameters of interest may be written as

$$F(x, \lambda, \beta) = \sigma(\beta(x - \lambda)), \quad (4)$$

where λ is the *threshold* and β the *sensitivity*. This formulation is equivalent to that in (2), with $\theta_1 = \beta$ and $\theta_0 = -\beta\lambda$. Figure 2a illustrates a sigmoidal psychometric function.

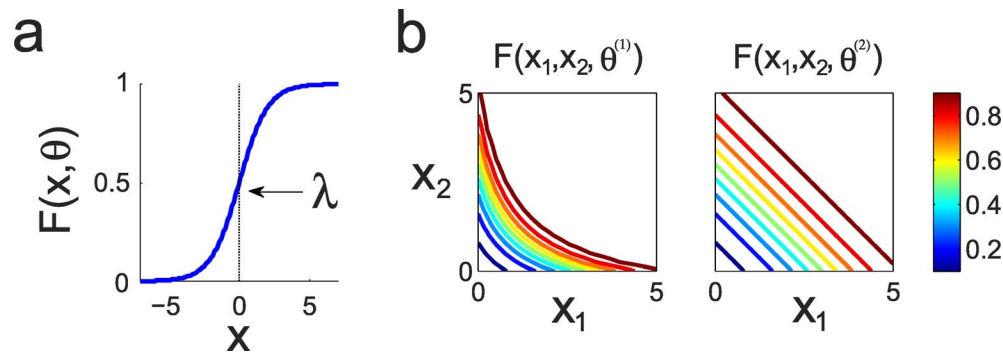


Figure 2. Examples of psychometric functions $F(\mathbf{x}, \theta)$ for one- and two-dimensional stimulus spaces. (a) A logistic psychometric function (Equation 2) with threshold $\lambda = 0$ and sensitivity $\beta = 1$. This function is one of several sigmoidal forms used in psychophysical research. (b) Level sets of the 2-D psychometric function (Equation 6) for two different values of the model parameter vector $\theta = (\theta_0, \theta_1, \theta_2, \theta_{12})^T$. Left: $\theta^{(1)} = (-3, 1, 1, 1)^T$. Right: $\theta^{(2)} = (-3, 1, 1, 0)^T$.

Multidimensional stimulus spaces

In many psychophysical experiments, it may be of interest to know how multiple stimulus parameters $\mathbf{x} = (x_1, \dots, x_N)^T$ interact to effect perception. For instance, recent investigations have considered how observers combine multiple cues like texture gradients and stereoscopic information for estimating the slant of a surface (Knill & Saunders, 2003; Hillis et al., 2004), while other investigations have considered how haptic and visual information are combined in performing reaching movements (Ernst & Banks, 2002). Other studies have considered how various cues are integrated when subjects perceive complex natural stimuli like occlusion boundaries or surfaces defined by multiple cues like texture, color, and luminance (Landy & Kojima, 2001; McGraw et al., 2003; Ing et al., 2010; DiMattina et al., 2012; Saarela & Landy, 2012). In the simplest case, cues are combined in a linear manner, and one can predict the response to simultaneous variations of multiple cues from the responses to each cue in isolation (Ernst & Banks, 2002; Knill & Saunders, 2003). However, several previous studies have demonstrated that sensory cues do not always combine linearly (Frome, Buck, & Boynton, 1981; Saunders & Knill, 2001; Zhou & Mel, 2008), making it necessary to covary several feature dimensions simultaneously to fully characterize the nature of their interactions.

Since the goal of this paper is to develop and analyze computational methods for multidimensional psychophysical experiments rather than to suggest particular models for sensory cue combination, I will mainly develop my implementations of the Ψ method in the context of generic multivariate logistic regression models (Bishop, 2006) which generalize the univariate model in Equation 2. These implementations can be readily extended to models of similar complexity which

are motivated by specific perceptual or neural hypotheses.

Generalizing the psychometric function

Generalization of the 1-D psychometric function (Equation 2) to two stimulus variables x_1 and x_2 which models the contribution of each individual variable and their possible multiplicative interactions may be written as

$$F(\mathbf{x}, \theta) = \sigma(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_{12} x_1 x_2 + \theta_{11} x_1^2 + \theta_{22} x_2^2). \quad (5)$$

In this paper, I will consider a simplified version of Equation 5 with zero diagonal terms, written

$$F(\mathbf{x}, \theta) = \sigma(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_{12} x_1 x_2). \quad (6)$$

An example 2-D psychometric function of the form in Equation 6 is illustrated in Figure 2b for two different parameter vectors θ . Generalizing Equation 6 to three stimulus variables is relatively straightforward, and in this case we have

$$F(\mathbf{x}, \theta) = \sigma(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \theta_{12} x_1 x_2 + \theta_{13} x_1 x_3 + \theta_{23} x_2 x_3 + \theta_{123} x_1 x_2 x_3). \quad (7)$$

It is important to note that these generalizations are certainly not the only possibilities for experiments involving multidimensional stimulus spaces, and other parameterization schemes have been proposed and used in experiments (Kujala & Lukka, 2006; Lesmes et al., 2010).

From Equations 6 and 7, we see that, when modeling all possible nondiagonal multiplicative interactions between stimulus parameters, increasing the dimensionality of the stimulus space causes the dimensionality of the parameter space to increase drastically. In

general, we find that for an N -dimensional stimulus space, the parameter-space dimensionality M will be

$$M = \sum_{k=0}^N \binom{N}{k}. \quad (8)$$

For the models in Equations 6 and 7, a simple reparameterization of the stimulus space allows us to write them as the linear multivariate logistic regression model encountered in statistics and machine learning (Bishop, 2006; Murphy, 2012). For instance, we may reparameterize the function in Equation 6 using the three variables $y_1 = x_1$, $y_2 = x_2$, and $y_3 = x_1 x_2$. The general form of this reparameterized model is

$$F(\mathbf{y}, \theta) = \sigma(\theta_0 + \sum_{i=1}^{M-1} \theta_i y_i), \quad (9)$$

where $\mathbf{y} = (y_1, \dots, y_{M-1})^T$ are the reparameterized stimuli, with M being the parameter-space dimensionality as given in Equation 8.

The estimation problem and adaptive stimulus design

As the number of model parameters increases, the amount of data needed to obtain a reliable estimate increases greatly, running up against the limitations posed by the amount of time available to work with an experimental subject. This *curse of dimensionality* (Bellman, 1961; Bishop, 2006) is quite familiar from studies using classification images (Mineault, Barthelme, & Pack, 2009; Murray, 2011), as well as sensory neurophysiology (Wu, David, & Gallant, 2006). Even in estimating the univariate psychometric function in Equation 2, a large amount of data must be collected in order to obtain sufficiently tight confidence intervals to see the effects of experimental manipulations on important psychometric function parameters like thresholds and slopes (Maloney, 1990; Wichmann & Hill, 2001). For this reason, numerous procedures have been developed to speed up the collection of psychophysical data using adaptive stimulus generation, the most popular of which is the Ψ method (Kontsevich & Tyler, 1999). These methods greatly decrease the number of stimuli needed to obtain reliable estimates of psychometric function parameters, at the cost of increased trial duration due to the need to iteratively compute the next stimulus.

Grid- Ψ

Method description

The most popular method for adaptive data collection in psychophysical research is the Ψ method, an information-theoretic algorithm which chooses stimuli with the goal of most rapidly reducing the

uncertainty about the parameter values as measured using the entropy of the posterior density (Kontsevich & Tyler, 1999). Although the usual grid-based implementation of the Ψ method is fast and effective for the standard two-parameter psychometric function (Equation 2) defined on a 1-D stimulus space (Kontsevich & Tyler, 1999; Prins, 2013a), it quickly becomes intractable for models defined in higher dimensional stimulus spaces with more parameters which must be estimated (Kujala & Lukka, 2006). I will refer to this standard implementation as Grid- Ψ .

The idea behind the Grid- Ψ method is that one maintains an evolving posterior density $\tilde{p}_n(\theta_k)$ defined on a discrete grid of potential model parameter values $S_\Theta = \{\theta_1, \dots, \theta_{N_\theta}\}$. In this paper I will use the notation $\tilde{p}_n(\theta_k)$ to denote a discrete density defined on my set of supports S_Θ (with $\sum_{k=1}^{N_\theta} \tilde{p}_n(\theta_k) = 1$), and the notation $p_n(\theta_k)$ to denote the original continuous density evaluated on the support θ_k . As new stimulus-response observations ($\mathbf{x}_{n+1}$, r_{n+1}) are obtained, this density is updated using Bayes's rule:

$$\tilde{p}_{n+1}(\theta_i) = \frac{1}{Z} p(r_{n+1} | \mathbf{x}_{n+1}, \theta_i) \tilde{p}_n(\theta_i), \quad (10)$$

where $Z = \sum_{i=1}^{N_\theta} p(r_{n+1} | \mathbf{x}_{n+1}, \theta_i) \tilde{p}_n(\theta_i)$ is a normalization constant. The uncertainty about the true parameter value is quantified using the Shannon entropy (Cover & Thomas, 2012) of the posterior distribution, given by the expression

$$H_n = - \sum_{i=1}^{N_\theta} \tilde{p}_n(\theta_i) \ln \tilde{p}_n(\theta_i). \quad (11)$$

At each step, the algorithm chooses the next stimulus $\mathbf{x}_{n+1}$ to present by finding the stimulus in a discrete set $X = \{\mathbf{x}_1, \dots, \mathbf{x}_{N_x}\}$ for which the expected value of the subsequent entropy H_{n+1} averaged over possible subject responses $r = 0$ or 1 is minimized. In other words, entropy is estimated by integrating out uncertainty about the true model parameters over the current posterior density $\tilde{p}_n(\theta_i)$ as well as uncertainty about the subject's responses $p(r | \mathbf{x})$. Mathematically, we write

$$\mathbf{x}_{n+1} = \arg \max_{\mathbf{x}} \langle -H_{n+1}(\mathbf{x}) \rangle_r, \quad (12)$$

where

$$\begin{aligned} \langle H_{n+1}(\mathbf{x}) \rangle_r &= H_{n+1}(\mathbf{x}, r=1) p(r=1 | \mathbf{x}) \\ &\quad + H_{n+1}(\mathbf{x}, r=0) p(r=0 | \mathbf{x}), \end{aligned} \quad (13)$$

with the calculation of $p(r | \mathbf{x})$ accomplished by integration over the current posterior. Efficient implementation of this procedure relies on a precomputed lookup table of the value of $p(r | \mathbf{x}_i, \theta_j)$ for $r = 0$ and 1 over all $\theta_j \in \Theta$ and $\mathbf{x}_i \in X$. A full description of the Ψ method and run-time analysis is given elsewhere (Kontsevich & Tyler, 1999; Kujala & Lukka, 2006).

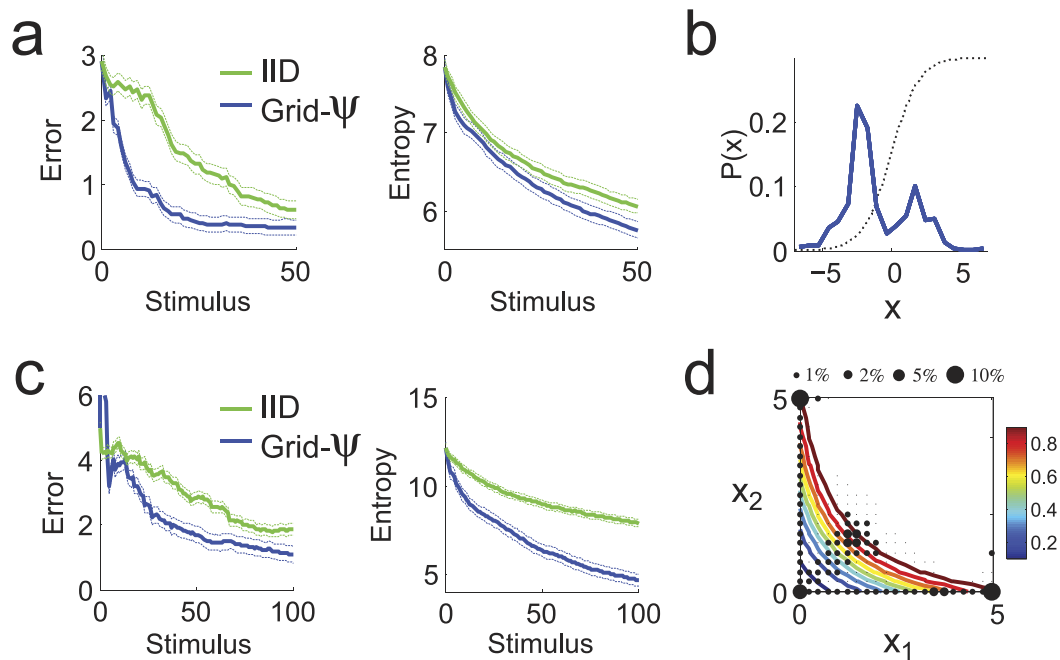


Figure 3. Performance of the standard Grid- Ψ method in simulated psychophysical experiments. (a) Left: Error E_n between current estimate and true observer parameters for uniform sampling (green) and the Grid- Ψ method (blue) averaged over 100 Monte Carlo trials. Thin dotted lines denote 95% confidence intervals. Right: Posterior entropy for both methods. (b) Placement of stimuli for the Grid- Ψ procedure (thick blue line). Overlaid is the true psychometric function, vertically scaled to $[0, 0.3]$ (dashed black line). (c) Same as (a) but for the two-dimensional psychometric function specified mathematically in Equation 6. (d) Stimulus placement for the 2-D Grid- Ψ procedure, overlaid on the contours of constant response probability. Black dots denote the unique stimuli presented, with the size of the dot proportional to how often the stimulus was presented. A compressive transformation ($\sqrt{u}$) is applied to enhance visibility of intermediate-sized dots, with percentage stimulus placements shown above the figure.

Numerical experiments

For the sake of comparison with the methods developed in this paper, I quantified the performance of the Grid- Ψ method for the case of $D_x = 1$ and $D_x = 2$ stimulus dimensions. I simulated psychophysical experiments in MATLAB on two different systems (3.30 GHz Intel Xeon-64 and 3.0 GHz Intel® i7-32) and quantified the improvements in parameter estimation due to the Grid- Ψ method as well as the per-trial running time. To quantify the estimation error on a trial-by-trial basis, we obtain a point estimate $\hat{\theta}_n$ of the parameter values by taking the expectation over the current posterior

$$\hat{\theta}_n = \int_{\Theta} \theta p_n(\theta) d\theta \quad (14)$$

and quantifying the error as $E_n = \|\hat{\theta}_n - \theta_T\|$. At the end of the experiment we also optimize the likelihood to attain the maximum likelihood (ML) estimate $\hat{\theta}_{ML}$, from which we obtain $E_{ML} = \|\hat{\theta}_{ML} - \theta_T\|$.

As we see in Figure 3a (left panel), in the 1-D case for a hypothetical observer with true parameters $\theta_T = (0, 1)^T$, the Grid- Ψ method (a special case of *optimal experimental design* or OED) greatly increases the accuracy of our parameter estimates when compared to

independent and identically distributed (IID) sampling from a discrete grid of evenly spaced points (i.e., the method of constant stimuli) and yields a posterior density with lower entropy (Figure 3a, right panel). Stimulus placement is shown in Figure 3b, with the blue curve indicating the empirical stimulus probabilities from all trials and all 100 Monte Carlo experiments. We see from this plot that the Grid- Ψ method concentrates its sampling near the endpoints of the linear region of the psychometric function (black dashed line, scaled to $[0, 0.3]$). These findings are consistent with the original study of the Ψ method (Kontsevich & Tyler, 1999), where stimulus placement was concentrated in a similar region of the psychometric function. Further analysis reveals that at the endpoints of the linear region, there is a large change in the probability of a correct response when the psychometric function parameters are varied, for both the sigmoidal form (Equation 2) and the form used by Kontsevich and Tyler (1999; see also Supplementary Figure S1).

In this example, I used $L_x = L_\theta = L = 51$ levels per stimulus and parameter dimension, with stimuli being chosen from a uniform grid on the interval $[-7, 7]$. To generate the grid of parameter values, I uniformly spaced $\log\beta \in [-1, 1]$, $\lambda \in [-2, 2]$ and computed $\theta_1 = \beta$,

$\theta_0 = -\beta\lambda$. When efficiently implemented using fully vectorized MATLAB/Octave code (available at <http://itech.fgcu.edu/faculty/cdimattina/>), the 1-D Grid- Ψ method is quite fast, taking less than 10 ms/trial (6.97 ± 0.18 ms, $N_{\text{trials}} = 5,000$) on a high-end workstation (Intel Xeon-64) and about 50 ms/trial (49.13 ± 1.31 ms, $N_{\text{trials}} = 5,000$) on a midrange system (Intel i7-32). Similar but more modest improvements were seen over the more restricted range of stimuli $[-5, 5]$, which more closely corresponds to the dynamic range of the 1-D sigmoid for this simulated observer in Figure 3a (Supplementary Figure S2). However, in general the appropriate dynamic range for a given observer may be unknown prior to the experiment, making it wise to err on the side of caution and to include at least a few stimuli along each tested dimension which will always be detected (or missed) by every subject.

Because of this exponential growth in run time as well as memory requirements of the Grid- Ψ method, implementing this procedure for the multivariate logistic regression model (Equation 6) with two stimulus dimensions becomes intractable at the same grid densities ($L = 51$) I used in the 1-D case. Using much less dense grids ($L = 21$) permitted implementation of this method, but it took nearly 4 s/trial (Intel Xeon-64 workstation) to generate the next stimulus (3.96 ± 0.013 s, $N_{\text{trials}} = 100$), making it far too slow for use in actual psychophysical experiments. In the 2-D implementation, I used a factorial grid of stimulus values $x_1, x_2 \in [0, 5]$ and defined a factorial grid of parameter values by uniformly spacing $\log\theta_1, \log\theta_2, \log\theta_{12} \in [-1, 1]$ and $\theta_0 = \lambda\beta$, where $\lambda \in [0, 4]$ and $\log\beta \in [-1, 1]$. As in the 1-D case, we obtain a substantial reduction in error (Figure 3c, left panel) and entropy (Figure 3c, right panel) with the Grid- Ψ procedure for a hypothetical observer having true parameters $\theta_T = (-3, 1, 1, 1)^T$. In this example, the true observer had a nonzero interaction term θ_{12} , which led to stimulus placement along the diagonal $x_1 = x_2$ of the stimulus space (Figure 3d) as well as along each of the individual stimulus axes. As with the 1-D case, the stimulus placement was located in regions of the stimulus space where there is a large change in the probability of correct response with respect to each of the psychometric function parameters (Supplementary Figure S3). This simple example nicely illustrates the necessity of simultaneously covarying stimulus parameters when there is the potential for nonlinear interactions. By contrast, for simulations on models similar to Equation 6 except without a nonlinear interaction term θ_{12} —i.e., $F(\mathbf{x}, \theta) = \sigma(\theta_0 + \theta_1 x_1 + \theta_2 x_2)$ —we find that stimulus placement is concentrated along the individual cardinal axes (Supplementary Figure S4). This validates the standard procedure of characterizing individual parameter dimensions separately when their interactions are linear (Hillis et al., 2004). I now consider three

alternative implementations of the Ψ procedure which are tractable in higher dimensions.

Prior- Ψ

Method description

The Ψ method requires approximating several integrals over a posterior density. In practice, these integrals are computed using a discrete *particle filter* (Carpenter, Clifford, & Fearnhead, 1999) which represents a continuous posterior density $p_n(\theta)$ by a density $\sum_{i=1}^{N_\theta} w_i^{(n)} \delta(\theta - \theta_i)$ defined on fixed supports $S_\Theta = \{\theta_1, \dots, \theta_{N_\theta}\}$, subject to the normalization constraint $\sum_{i=1}^{N_\theta} w_i^{(n)} = 1$. One can approximate expectations with respect to $p_n(\theta)$ using the expression

$$\int_{\Theta} f(\theta) p_n(\theta) d\theta \approx \sum_{i=1}^{N_\theta} w_i^{(n)} f(\theta_i), \quad (15)$$

where $w_i^{(n)}$ is the importance weight associated with each particle (Carpenter et al., 1999; Arulampalam, Maskell, Gordon, & Clapp, 2002). In my implementation, the supports S_Θ are sampled from the prior $p_0(\theta)$ and fixed through the experiment ($\theta_i^{(n)} = \theta_i$ for all n). The importance weights are initially set equal to $1/N_\theta$ and evolve using sequential importance sampling, where the importance function $\pi_0(\theta^{(n)}|\theta^{(1:n-1)}, r_{1:n})$ and state transition function $p(\theta^{(n)}|\theta^{(1:n-1)})$ are both equal to the prior $p_0(\theta)$, leading to the simplified importance-weight update rule:

$$\begin{aligned} w_i^{(n+1)} &\propto w_i^{(n)} \frac{p(r_{n+1}|\mathbf{x}_{n+1}, \theta_i^{(n+1)}) p(\theta_i^{(n+1)}|\theta_i^{(1:n)})}{\pi_0(\theta_i^{(n+1)}|\theta_i^{(1:n)}, r_{1:(n+1)})} \\ &= w_i^{(n)} p(r_{n+1}|\mathbf{x}_{n+1}, \theta_i). \end{aligned} \quad (16)$$

Note that Equation 16 is simply sequential Bayesian updating of the discrete posterior $w_i^{(n)} = \tilde{p}_n(\theta_i)$, with $\tilde{p}_0(\theta_i) = w_i^{(0)} = 1/N_\theta$.

A previous study (Kujala & Lukka, 2006) suggested that one potential approach to increasing the tractability of the Ψ method is to abandon the factorial, grid-based representation of the posterior density used by Kontsevich and Tyler (1999) and instead represent the posterior on a tractable number N_θ of supports sampled from some prior density $p_0(\theta)$. This naturally gives rise to the question of how to specify an appropriate prior distribution. One recent study (Kim, Pitt, Lu, Steyvers, & Myung, 2014) suggests a principled way to specify priors, namely via hierarchical Bayesian modeling where the results of previous experiments are used to estimate a hyper-prior which can be used to define a prior for subsequent experiments. In the present study, I am interested in the effects of using different sampling strategies (adaptive and nonadaptive) while controlling for the prior shape, and I do not systematically consider

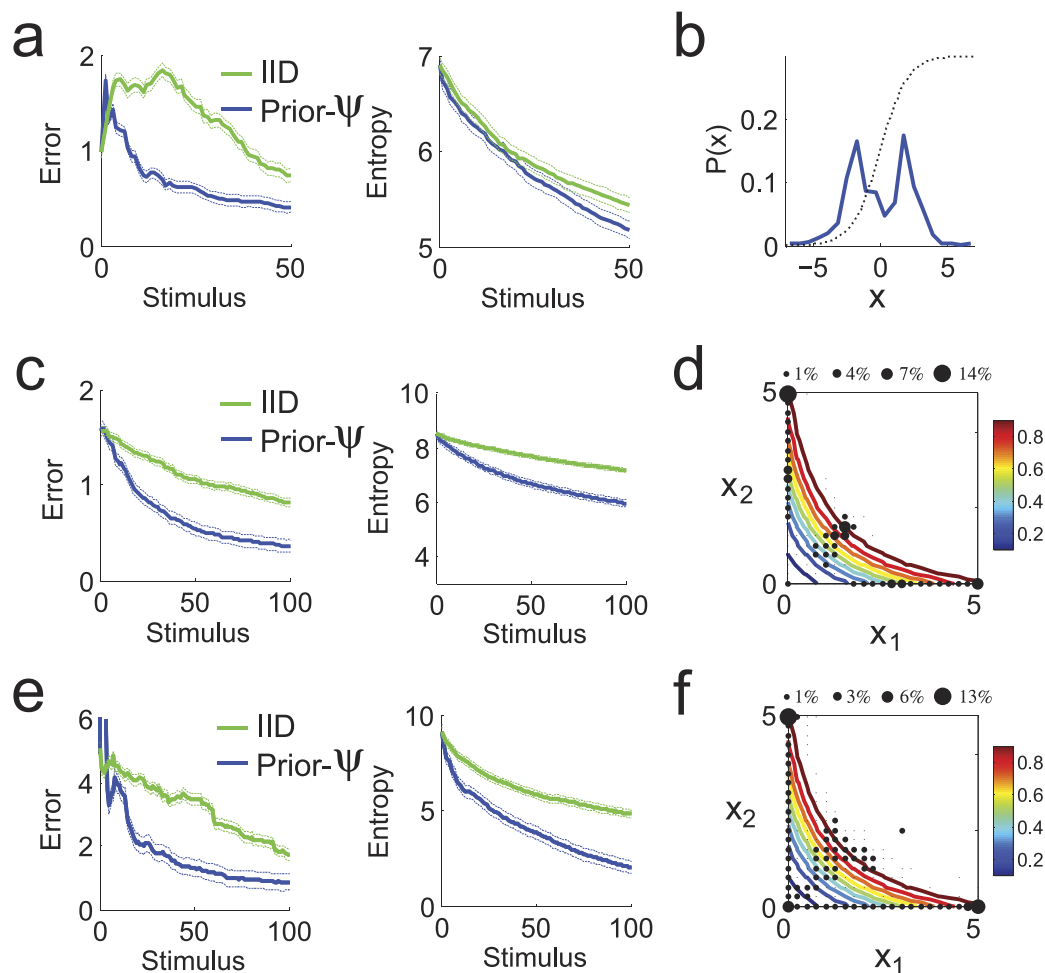


Figure 4. Performance of the of the Prior- Ψ method. (a–d) For a Gaussian prior, with the same organization as Figure 2. (e–f) For a uniform prior, with the same organization as Figure 2c and d.

the problem of prior specification or possible effects of prior misspecification.

Although this idea of using a set of particles sampled from a prior density has been proposed previously in the psychometric literature, it has only been fully implemented and analyzed for a specialized two-dimensional psychometric function which parameterizes 2-D thresholds as ellipses (Kujala & Lukka, 2006). In contrast, my treatment here is much more general, as multivariate logistic regression (Equation 9) is a generic machine-learning model which is applicable to many possible types of experiment (Bishop, 2006). I refer to this as the Prior- Ψ method.

Numerical experiments

We see from Figure 4 that using a tractable number of particles sampled from an informative prior density manages to reproduce the main results seen in Figure 3 for the Grid- Ψ method. In the 1-D example shown in Figure 4a and b, I used $N_\theta = 1,000$ particles, and in the 2-D examples (Figure 4c through f) I used $N_\theta = 5,000$ or

10,000. For comparison, in the Grid- Ψ method for 2-D with $L = 21$ levels per stimulus dimension, we have $N_\theta = L^4 = 194,481$ particles. In one 2-D example (Figure 4c, d), I used a Gaussian prior with σ in each dimension equal to one half the upper and lower bounds of the grids used in the Grid- Ψ example. This guaranteed that the set of particles sampled approximately the same range of parameter values as before. Similar results were obtained using the uniform prior implicitly assumed in the Grid- Ψ method, from which the same numbers of particles were sampled (Figure 4e, f).

Figure 5 shows average stimulus generation times for the 2-D experiment and median final estimation error E_n as a function of the number of particles N_θ used to represent the posterior for Prior- Ψ and Grid- Ψ (black diamond). Supplementary Figure S5 shows the median final error E_{ML} . We see from Figure 5 (left panel) that one can certainly increase the speed of the implementation by reducing the number of particles used to represent the posterior, but this may reduce the accuracy of the final parameter estimates (Figure 5, right panel; Supplementary Figure S5). On the 32-bit

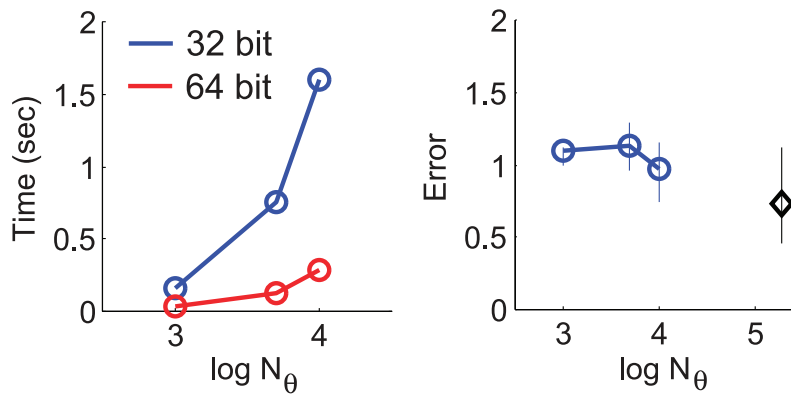


Figure 5. Left: Stimulus selection times for Prior-Ψ as a function of the number of particles N_θ used to represent the posterior. Right: Median final error E_n as a function of N_θ . Blue circles indicate Prior-Ψ, and black diamonds Grid-Ψ. Bars indicate 25th through 75th percentiles (100 Monte Carlo trials). In this example I used the uniform prior implicit in the standard implementation of the Ψ method.

system, implementing Prior-Ψ can potentially be slow in cases where large numbers of particles are needed ($N_\theta = 5,000$: 737 ± 31 ms; $N_\theta = 10,000$: 1.60 ± 0.39 s, $N_{\text{trials}} = 10,000$). Although stimulus selection times of nearly 1 s may be acceptable for many experiments, over thousands of trials this overhead can add tens of minutes to the experiment duration. Therefore, it is of great interest to develop faster implementations of the Ψ method, especially for generalizations into even more stimulus dimensions where more particles are needed to accurately represent the posterior density.

In my implementation of Prior-Ψ, I do not implement the full version implemented by Kujala and Lukka (2006), which included after each trial a step where a new set of particles was sampled from the continuous posterior using Markov-chain Monte Carlo (MCMC) methods (Gilks, 2005). In many situations, it may be useful to update the particle filter, since one well-known limitation of particle-filter approximations to an evolving posterior density is the fact that as the experiment progresses, fewer particles θ_i have probability $\tilde{p}_n(\theta_i)$ substantially greater than 0 (Bengtsson, Bickel, & Li, 2008; Bickel, Li, & Bengtsson, 2008; Snyder, Bengtsson, Bickel, & Anderson, 2008). Investigators in computational statistics (Arulampalam et al., 2002) have developed a measure to quantify the number N_{eff} of effective particles in a particle filter, given by the expression

$$N_{\text{eff}} = \frac{1}{\sum_{i=1}^{N_\theta} \tilde{p}_n(\theta_i)^2}. \quad (17)$$

Supplementary Figure S6 illustrates how N_{eff} quickly decreases during the course of the experiments presented in Figure 4. Although my examples did not require resampling for accurate estimation, for some problems it may be useful. The trade-off to consider is

whether the improvement in estimation accuracy justifies the additional time needed for the MCMC resampling step, which takes longer as the experiment progresses, since the likelihood function has more terms. This decision is clearly dependent on the particular problem.

Lookup-Ψ

Method description

Another method for potentially speeding up adaptive design of psychophysical experiments originates from a set of ideas developed in the statistics literature, namely the theory of optimal experimental design (Chaloner & Verdinelli, 1995; Atkinson et al., 2007). OED chooses a set of observation points $d = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m\}$ known as a *design*, which will optimize some utility function $U(d)$ specifying some desired experimental goal, for instance prediction of new observations (O'Hagan & Kingman, 1978), model comparison (Cavagnaro et al., 2010; Cavagnaro, Pitt, & Myung, 2011), or accurate estimation of model parameters (Lewi et al., 2009). For parameter-estimation problems, a valid design must include at least as many points as there are parameter dimensions: For instance, for the standard 1-D psychometric function shown in Figure 1a (which has two parameters), a valid design $d = \{\mathbf{x}_1, \mathbf{x}_2\}$ for estimation must contain two unique stimuli. In a Bayesian setting, we may formally state the design problem as finding

$$d^* = \arg \max_d U(d), \quad (18)$$

where

$$U(d) = \int \int u(d, \theta, \mathbf{y}) p(\mathbf{y}|d, \theta) p_0(\theta) d\mathbf{y} d\theta. \quad (19)$$

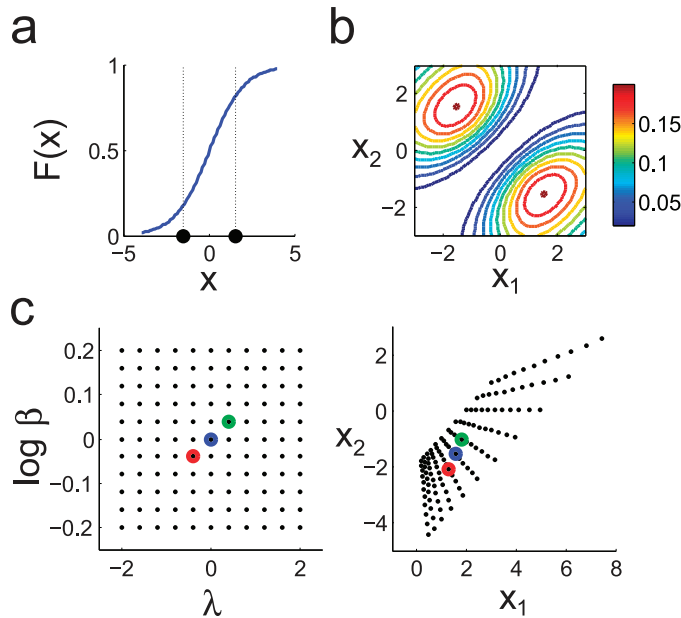


Figure 6. Illustration of the D-optimal designs for estimating the parameters of the 1-D psychometric function. (a) D-optimal design $d = \{x_1, x_2\}$ (black dots) for true parameters $(\lambda, \beta) = (0, 1)$ for the 1-D psychometric function (Equation 2). (b) Conditional expected utility $U(d|\theta)$ for the space of two-element designs $d = \{x_1, x_2\}$ for the function in (a). (c) Mapping between psychometric function parameters and D-optimal designs. Note that nearby parameter values are mapped to similar designs (red, green, and blue circles).

Integrating over the possible observations $\mathbf{y}$ conditioned on d, θ , we obtain the expression

$$U(d) = \int U(d|\theta) p_0(\theta) d\theta, \quad (20)$$

where the quantity $U(d|\theta)$ is the *conditional expected utility* of the design d . In general, exact evaluation of the integrals in Equations 19 and 20 is not analytically tractable and typically requires use of numerical Monte Carlo methods (Chaloner & Verdinelli, 1995). OED can also be implemented sequentially using a greedy algorithm where only a single stimulus is chosen on each trial to maximize the expected utility, and in fact many of the information-theoretic stimulus design methods in the neuroscience and psychology literature are simply special cases of sequential optimal design with expected posterior entropy as the utility function (Kontsevich & Tyler, 1999; Lewi et al., 2009; DiMattina & Zhang, 2011; Cavagnaro et al., 2010).

The form of Equation 20 suggests a simple heuristic strategy which I call *Lookup-Ψ* for sequential data collection in the case where one is representing the posterior density on a finite set of supports $S_\Theta = \{\theta_1, \theta_2, \dots, \theta_{N_\theta}\}$ and there are a finite number of designs $D = \{d_1, d_2, \dots, d_m\}$ we may present. Assuming that we have precomputed the conditional expected utility $U(d|\theta_i)$ of each design d_j for all θ_i , we can then

approximate the expected utility of design d_j for trial $n + 1$ given our current discrete posterior $\tilde{p}_n(\theta_i)$ by calculating

$$U_{n+1}(d_j) = \sum_{i=1}^{N_\theta} U(d_j|\theta_i) \tilde{p}_n(\theta_i). \quad (21)$$

This expected utility of each of our possible designs in D may be computed using a simple matrix multiplication

$$\mathbf{q}_{n+1} = \mathbf{U} \mathbf{p}_n, \quad (22)$$

where $\mathbf{U}_{ji} = U(d_j|\theta_i)$, $[\mathbf{p}_n]_i = \tilde{p}_n(\theta_i)$, and $[\mathbf{q}_{n+1}]_j = U_{n+1}(d_j)$. At each iteration, we present the design which maximizes the expected utility, i.e., the design corresponding to the largest component of $\mathbf{q}_{n+1}$. In order to ensure that there are designs in our set D which are optimal for the possible states of nature in S_Θ , prior to the experiment we precompute for each of the $\theta_i \in S_\Theta$ the optimal design d_i^* given by

$$d_i^* = \arg \max_d U(d|\theta_i). \quad (23)$$

Computing the optimal design for each possible state of nature yields an $N_\theta \times N_\theta$ matrix $\mathbf{U}$. However, since a design which is optimal for a given $\theta_i \in S_\Theta$ is also nearly optimal for nearby $\theta_k \in \Theta$, we can reduce the dimensionality of the matrix $\mathbf{U}$ and thereby speed up the multiplication in Equation 21 by only including designs in D for a subset of N_d points $\theta_i \in \Theta$.

It is important to note that the *Lookup-Ψ* method differs in several critical respects from previously proposed methods. In particular, instead of choosing a new stimulus on every trial like the other methods, it chooses a set of $K = \dim(\theta)$ stimuli (i.e., a design $d \in D$) every K trials. Also, the term “lookup” should not be confused with the precomputation step in the standard Ψ technique, which simply precomputes $p(r|\mathbf{x}_i, \theta_j)$ for all stimuli $\mathbf{x}_i$ and supports θ_j . In this context, “lookup” refers to the fact that for each design in our set, we precompute the expected utility of that design for each of the supports in our discrete approximation of the posterior and store the value in a matrix $\mathbf{U}$ which remains fixed throughout the experiment. Since the matrix $\mathbf{U}$ is defined on the fixed set of supports specified prior to the experiment, this method may not be as readily amenable to MCMC resampling as *Prior-Ψ*, due to the need to also update the design set D and the utility matrix $\mathbf{U}$ along with the particles.

In my implementation of the method, I used as my utility function D-optimality (Atkinson et al., 2007), which chooses stimuli to maximize the determinant of the Fisher information matrix. However, many other choices are possible and commonly used in the OED literature (Chaloner & Verdinelli, 1995; Atkinson et al., 2007). Figure 6 concretely illustrates the ideas I have

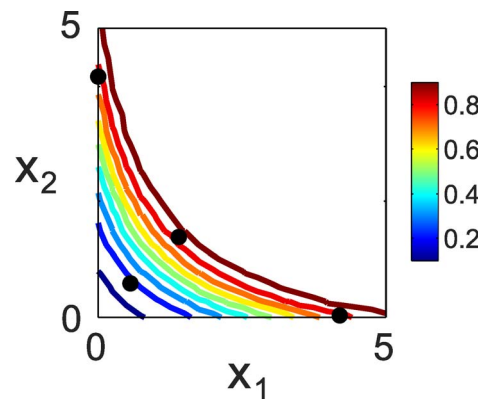


Figure 7. The four element D-optimal design (black dots) for estimating the model of Equation 6 with true parameters $\theta_T = (-3, 1, 1, 1)^T$.

described for the case of the 1-D psychometric function (Equation 2). In Figure 6a we see that the D-optimal design d^* for estimating the true parameters $\theta_T = (0, 1)$ —equivalent to $\lambda = 0$, $\beta = 1$ —is given by the stimuli $x_1, x_2 = \pm 1.54$. These stimuli are located at the regions of the psychometric curve where the change in the observer's correct response probability with respect to each of the parameters is large (Supplementary Figure S1). We see that these two stimuli are where all of my implementations of the 1-D Ψ method concentrate much of their sampling (Figures 3b and 4b), eventually alternating between these two stimuli as the experiment progresses. This fact makes sense in light of the fact that asymptotically, the posterior density becomes well described by a Gaussian (Kay, 1993), at which point minimizing the entropy is equivalent to maximizing the Fisher information determinant (Atkinson et al., 2007). Indeed, in the original implementation of the Ψ method, as the experiment progressed the stimulus presentations also alternated between two stimuli located near regions of their psychometric function where the response probability had a large change with respect to the function parameters (Kontsevich & Tyler, 1999).

Although this design $d = \{x_1, x_2\} = \pm 1.54$ is optimal for estimating parameters $\lambda = 0$ and $\beta = 1$, we see from Figure 6b that nearby designs also have high utility. This near optimality of nearby designs makes it possible to compute optimal designs for only a subset of the N_θ particles used to represent the posterior density, thereby reducing the space and time requirements of the lookup-table method. Figure 6c illustrates for a grid of different parameter values (left panel) the optimal designs (right panel) for those values. We see that nearby parameter values map to nearby optimal designs, which is consistent with the fact that the mapping between parameter and design space is smooth and continuous. A derivation of the D-optimal design for Equation 2 is shown in Figure 6a, and an

analytical approximation to this design is given in the Appendix. Similarly, the numerically optimized D-optimal design for the 2-D psychometric function (Equation 6) is shown in Figure 7. Note that it concentrates design points in regions of the psychometric function where the change in response probability with respect to each of the parameters is high (Supplementary Figure S3). However, it is important to remember that the D-optimal design for a set of parameters is not simply the concatenation of D-optimal designs for each parameter individually. Therefore, the D-optimal design in Figure 7 is not identical to the locations of maxima in Supplementary Figure S3.

Numerical experiments

I implemented the Lookup- Ψ method for 1-D and 2-D psychometric estimation, obtaining improvements in estimation accuracy and stimulus placement similar to those seen in my other implementations (Figure 8). The Lookup- Ψ method was extremely fast, taking only about 1 ms (0.96 ± 0.2 ms, $N_{\text{trials}} = 2,500$) on the 64-bit Xeon and 3 ms (2.52 ± 0.37 ms, $N_{\text{trials}} = 2,500$) on the 32-bit i7 to generate the next (four-stimulus) design for the 2-D psychometric function with $N_\theta = 5,000$ particles and $N_d = 1,250$ designs. By contrast, the Prior- Ψ procedure with the same number of particles used to represent the posterior ($N_\theta = 5,000$) took nearly 1 s/stimulus (737 ± 31 ms, $N_{\text{trials}} = 10,000$) on the 32-bit system for estimating this same 2-D psychometric function. A direct comparison of methods with $N_\theta = 5,000$ particles sampled from the same Gaussian prior demonstrates that the Lookup- Ψ method offers a tremendous speedup over the Prior- Ψ procedure ($300\times$ on the i7) without sacrificing final accuracy as measured by Equation 14 (Prior- Ψ error = 0.360 ± 0.239 ; Lookup- Ψ error = 0.369 ± 0.231 , $N_{\text{mc}} = 100$).

Laplace- Ψ

Method description

Representing high-dimensional prior and posterior densities poses a formidable computational challenge for Bayesian statistical learning. One approach to efficiently representing the posterior used in previous studies of adaptive stimulus design for neurophysiology is to make use of the Laplace approximation (Bishop, 2006; Lewi et al., 2009). I demonstrate here how this method can be fruitfully be extended to the Ψ procedure.

The basic idea of the Laplace approximation is to represent the evolving posterior density $p_n(\theta)$ as a Gaussian centered on the posterior mode μ_n . As new observations $(\mathbf{x}_{n+1}, r_{n+1})$ are obtained, the mode and

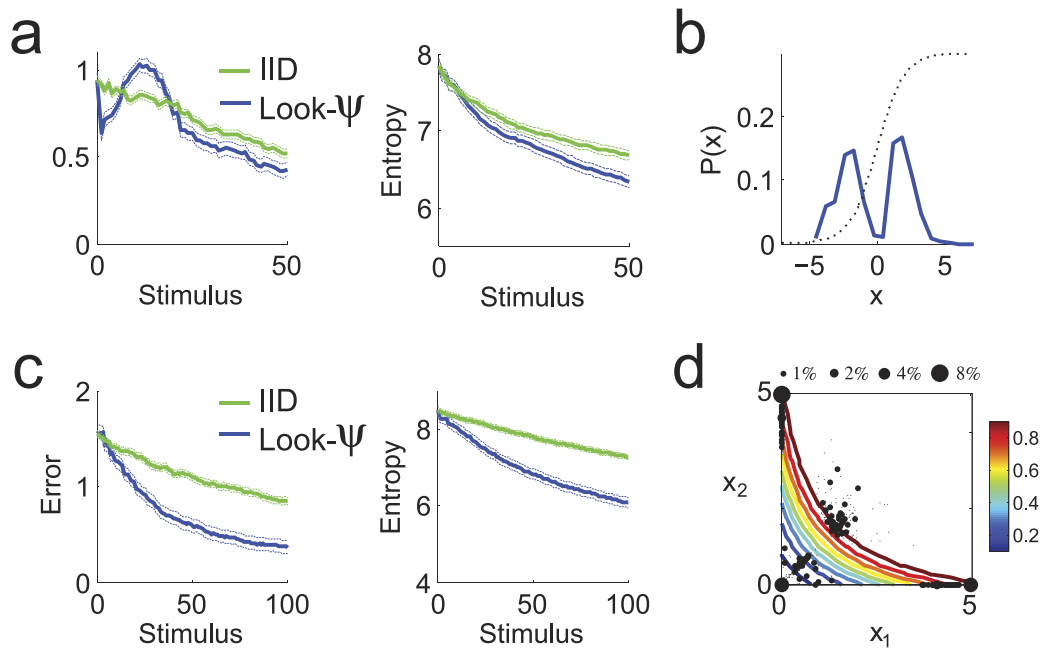


Figure 8. Results for the Lookup- Ψ method for 1-D and 2-D psychometric functions. Organization is the same as in Figure 3.

covariance estimates are updated until the mode converges to a final estimate with a sufficiently small covariance. In the case where the system response model $p(r|\mathbf{x}, \boldsymbol{\theta})$ is Gaussian, one may analytically compute the new mode $\boldsymbol{\mu}_{n+1}$ and new covariance $\boldsymbol{\Sigma}_{n+1}$ recursively using the Kalman-filter update formulas (Kay, 1993). However, when the response model is not Gaussian or cannot be reasonably approximated as one (for instance, the psychophysical model in Equation 1), the new posterior mode $\boldsymbol{\mu}_{n+1}$ cannot be computed analytically and must be found using numerical optimization. However, since a single observation generally does not drastically change the location of the posterior mode, this optimization is quite fast. Given the new mode $\boldsymbol{\mu}_{n+1}$, computation of the covariance $\boldsymbol{\Sigma}_{n+1}$ is straightforward using the formulas presented in this section.

Mathematically, the Laplace approximation to the current posterior density is obtained via the second-order Taylor expansion around the current mode $\boldsymbol{\mu}_n$ of the log-posterior:

$$\ln p_n(\boldsymbol{\theta}|D_{1:n}) \approx \ln p_n(\boldsymbol{\mu}_n|D_{1:n}) - \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\mu}_n)^T \boldsymbol{\Sigma}_n^{-1}(\boldsymbol{\theta} - \boldsymbol{\mu}_n), \quad (24)$$

where

$$\boldsymbol{\Sigma}_n^{-1} = -\nabla \nabla_{\boldsymbol{\theta}} \ln p_n(\boldsymbol{\mu}_n|D_{1:n}). \quad (25)$$

and $D_{1:n} = \{(\mathbf{x}_1, r_1), (\mathbf{x}_2, r_2), \dots, (\mathbf{x}_n, r_n)\}$ is all of the stimulus-response data collected during the first n trials. Exponentiating both sides of Equation 24 yields a Gaussian approximation p_n centered at $\boldsymbol{\mu}_n$ with

covariance $\boldsymbol{\Sigma}_n$, with the entropy of this Gaussian density given by

$$H_n = \frac{1}{2} \ln \det(\boldsymbol{\Sigma}_n). \quad (26)$$

In choosing the next stimulus $\mathbf{x}_{n+1}$, the Ψ method minimizes the expected entropy of the subsequent Gaussian posterior $p_{n+1}(\boldsymbol{\theta}|D_{1:n}, r_{n+1}, \mathbf{x}_{n+1})$, where $r_{n+1} \in \{0, 1\}$ is the (unknown) subject response for trial $n+1$.

The calculation of $p(r|\mathbf{x}) = \int p(r|\mathbf{x}, \boldsymbol{\theta}) p_n(\boldsymbol{\theta}) d\boldsymbol{\theta}$ is straightforward and can be accomplished by Monte Carlo integration. From Equation 26 we see that computing the entropies $H_{n+1}(\mathbf{x}, r)$ simply amounts to calculating the two possible determinants of $\boldsymbol{\Sigma}_{n+1}$, which we find using

$$\boldsymbol{\Sigma}_{n+1}^{-1} = -\nabla \nabla_{\boldsymbol{\theta}} \ln p(r|\mathbf{x}, \boldsymbol{\mu}_n) + \boldsymbol{\Sigma}_n^{-1} \quad (27)$$

for the cases where $r=0$ and $r=1$. Note that we evaluate Equation 27 using the current posterior mode $\boldsymbol{\mu}_n$, which is a reasonable approximation assuming that successive posterior modes are nearby. Applying this method to the generic multivariate regression model (Equation 9), we readily compute

$$-\nabla \nabla_{\boldsymbol{\theta}} \ln p(r|\mathbf{x}, \boldsymbol{\mu}_n) = \sigma'(r) \mathbf{x}^T \boldsymbol{\mu}_n \mathbf{x} \mathbf{x}^T. \quad (28)$$

One question which arises when using the Laplace approximation is whether it is reasonable to approximate the posterior density by a single Gaussian bump, as such a representation would be inappropriate for a density with multiple peaks. However, it is simple to show that for the logistic regression model, the likelihood is log-concave (Pratt, 1981), and hence with

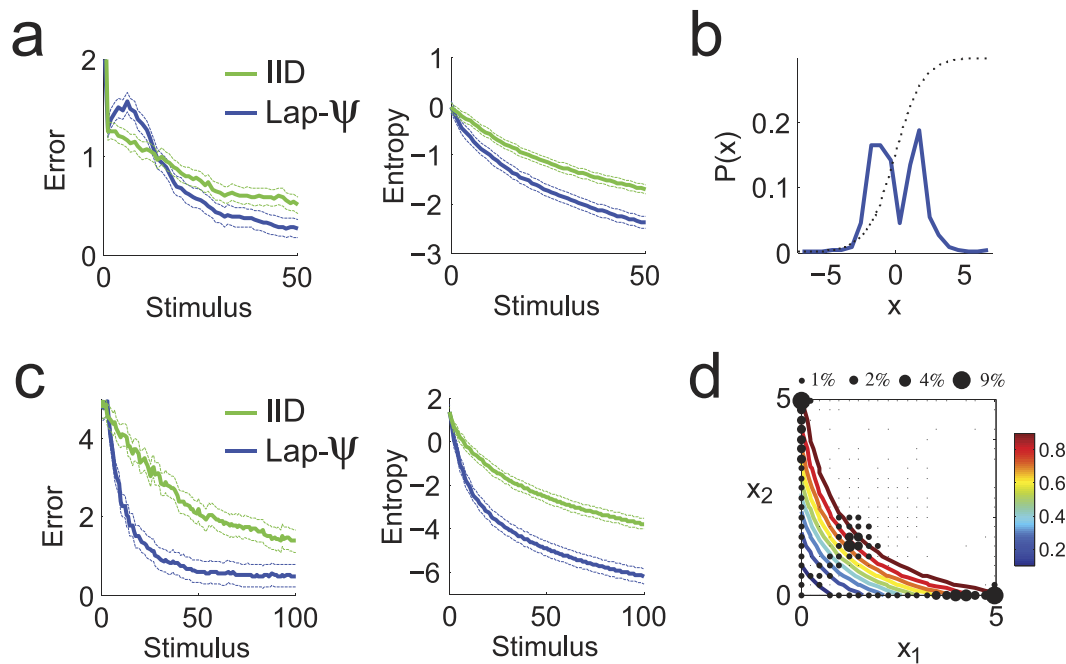


Figure 9. Results for the Laplace- Ψ method for 1-D and 2-D psychometric functions. Organization is the same as in Figure 3.

a log-concave prior (i.e., a Gaussian, Laplace, or flat prior) the posterior cannot have more than one maximum. For psychometric functions where the likelihood does not enjoy log-concavity (for instance, those defined using multiple-layer neural networks), the Laplace- Ψ method may be inappropriate, and other methods like particle filters (Carpenter et al., 1999) or Gaussian-sum approximations of the posterior (Alspach & Sorenson, 1972; DiMattina & Zhang, 2011) may be more suitable. Another potential issue is that this method assumes a small change between successive posterior modes, which is reasonable if enough data have been collected but is unlikely to be the case early on in the experiment. Nevertheless, I found the method to work quite well in my examples, and previous work has successfully applied the Laplace approximation method to estimating neural receptive fields (Lewi et al., 2009).

Numerical experiments

I implemented the Laplace- Ψ algorithm for 1-D and 2-D psychometric functions (Equations 2 and 6) to permit comparisons with the methods presented previously (Figure 9). I used the same true parameters θ_T and discrete stimulus search grid as the other examples ($L = 21$ evenly spaced stimuli on $[0, 5]$ in the 2-D case), and defined my initial prior with $\mu_0 = (-1, 0.5, 0.5, 0.5)^T$ and $\Sigma_0 = 2 \cdot \mathbf{I}$. I found that the Laplace- Ψ procedure was extremely fast, taking only about 20 ms (18.83 ± 1.88 ms, $N_{\text{trials}} = 10,000$) to choose the next stimulus on the 32-bit Intel i7 for the 2-D psychometric

function. In my implementations, due to the relatively low dimensionality of the parameter space (4–8 dimensions), I made use of the MATLAB/Octave command `fminunc.m` (supplied with gradient and starting from μ_n) to update the posterior mode and found this to be fast (15.69 ± 2.91 ms, $N = 100$) for the highest dimensional examples I analyzed.

Efficiency of OED methods

All of the implementations discussed so far provide a tremendous advantage over the method of constant stimuli (IID sampling from a grid) in terms of greatly reducing the number of trials needed to reach a desired value of the expected square error or entropy. We can define the trial speedup factor for each method as follows: Given a final expected squared error E_{iid} (Equation 14) obtained by IID stimulus presentation after N trials, we define the trial speedup factor $S = N/N_{oed}$, where N_{oed} is the number of trials needed on average to reach the criterion error E_{iid} . This factor tells us how many times faster OED methods are than IID methods for attaining the same accuracy. Supplementary Figure S7 plots this speedup factor for all of the methods for the case of the 2-D psychometric function (Equation 6). We see that for this particular example, each of these methods is about 3.5–4.5 times more efficient than IID sampling in terms of the savings in number of stimulus presentations. Clearly, the degree benefit obtained by adaptive stimulus optimization methods will in general be dependent on the particular problem. Therefore, the

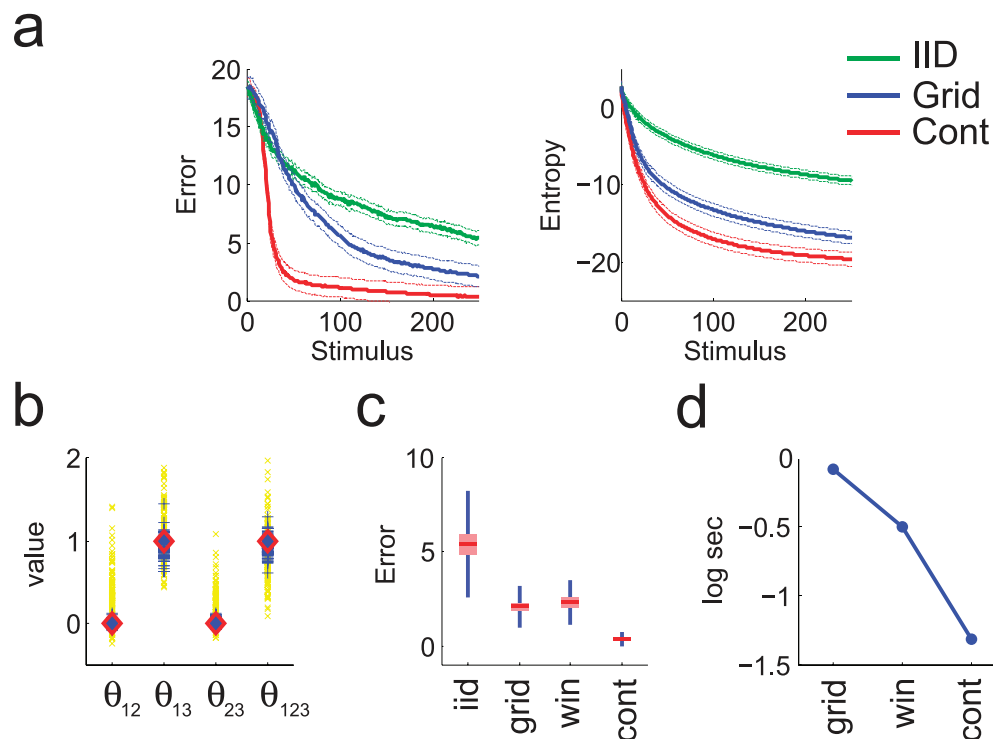


Figure 10. Stimulus optimization in discrete and continuous spaces for a psychometric model with a 3-D stimulus space. (a) Optimizing stimuli in a continuous 3-D space (red curve) results in more accurate parameter estimates with less posterior entropy when compared with optimization on a grid (blue). (b) Accurate estimation of interaction terms with continuous stimulus optimization (blue symbols) compared with random stimuli (yellow symbols). (c) Final estimation error (Equation 14) for various stimulus optimization strategies. “Win” stands for an approach where the stimulus set is periodically winnowed to eliminate stimuli of low expected utility. We see that all OED methods vastly improve accuracy, and continuous stimulus optimization yields the most accurate estimates. (d) Time for various stimulus-space searching methods, averaged across the experiment. We see that continuous stimulus optimization is the most efficient in this example.

results presented here in the context of a few specific examples provide an existence proof that such methods may be useful in some situations.

Nonlinear interactions of multiple cues

Investigating nonlinear cue combination

For many sensory quantities, there are often several cues which can give estimates of those quantities. For instance, one can estimate the tilt of a surface in depth from both texture and stereoscopic cues (Knill & Saunders, 2003; Hillis et al., 2004). Similarly, in natural vision multiple cues can be used for segmenting surfaces or detecting occlusions (Figure 1), including texture, luminance, and color (Konishi, Yuille, Coughlan, & Zhu, 2003; Martin, Fowlkes, & Malik, 2004; Ing et al., 2010; DiMattina et al., 2012). A large body of work in sensory psychophysics has considered what the optimal strategy is for combining multiple cues when estimating a sensory quantity. However, most of this work has focused on testing if subjects combine cues linearly in an optimal manner (Trommershauser et al., 2011). In

general, there is no reason why perceptual cues should necessarily be combined in a linear manner, and in fact recent computational work suggests that linear cue combination is actually suboptimal for integrating information from different color channels (Zhou & Mell, 2008). Studying nonlinear cue combination for complex stimuli defined by multiple parameters necessitates the use of high-dimensional models like Equation 9.

In order to investigate the potential utility of the Laplace- Ψ method for 3-D cue-combination experiments, I considered estimating the true parameters $\theta_T = (-4, 1, 1, 0.5, 0, 1, 0, 1)^T$ of the eight-parameter model (Equation 6) defined on three-dimensional stimuli $\mathbf{x} = (x_1, x_2, x_3)^T$. In this hypothetical example, stimulus feature x_3 interacts multiplicatively with stimulus feature x_1 , and therefore it is inappropriate to assume linear cue combination. Our stimulus space is a discrete 3-D grid with $L = 21$ levels uniformly spaced on $[0, 4]$, for a total of $(21)^3 = 9,261$ possible stimuli. Our Gaussian prior had mean $\mu_0 = (0, 0.5, 0.5, 0.5, 0, 0, 0, 0)^T$ and covariance $\Sigma_0 = 2 \cdot \mathbf{I}$.

We see from Figure 10a that there is a substantial improvement in performance when compared to IID

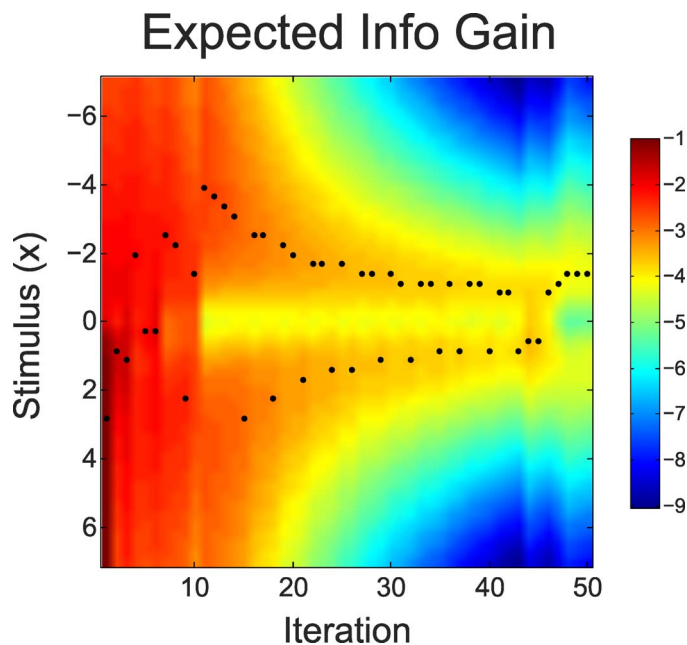


Figure 11. Evolution of the expected information gain (plotted on log scale) for a run of the 1-D Grid- Ψ method. Note that as the experiment progresses, fewer stimuli have substantial expected information gain.

sampling, and furthermore the nonlinear interaction terms are more accurately recovered (Figure 10b). We find that as in the 2-D simulations with a nonlinear interaction term θ_{12} (Figure 3), many of the stimuli chosen by the Ψ procedure lie on the diagonals of the stimulus space (Supplementary Figure S8). This demonstrates the crucial importance of covarying stimulus features when their perceptual interactions are nonlinear. I found that when searching over the full grid of $N = 9,261$ stimuli for the stimulus maximizing Equation 13, stimulus selection times were somewhat slow (Xeon-64: 827 ± 5 ms, $N_{\text{trials}} = 25,000$) due to the large number of function evaluations needed. I now demonstrate two possible ways to speed up the stimulus selection times.

Speeding up implementation

As one moves into higher dimensions, it may not be feasible to evaluate the expected information gain (Equation 13) on a grid of all possible combinations of stimulus values, as the function evaluations grow exponentially as L^D , where L is the number of levels per dimension and D is the dimensionality. Therefore, one may wish to reduce the number of function evaluations by either optimizing Equation 13 on a continuous stimulus space or periodically winnowing the stimulus space so that only stimuli which have had high utility over past trials are retained. The justification for this winnowing of the stimulus space is that as one learns

more about the true parameters of the system, fewer stimuli are potentially useful for refining one's estimate. The expected information gain (normalized by the current entropy) is illustrated in Figure 11 for the 1-D psychometric function (Equation 2), and we see that as the experiment progresses, fewer stimuli have high expected utility, with the algorithm eventually alternating between the two points (± 1.54) which comprise the D-optimal design for the true system parameters (Figure 6a).

We see in Figure 10 that optimizing over a continuous stimulus space and periodic winnowing by rank-ordering stimuli's expected information gain (eliminating the bottom 25% of stimuli every 25 trials) both offer substantial reductions in stimulus selection time compared to searching over the 3-D grid of stimuli as in the standard implementation of the Ψ method (Figure 10d). Furthermore, we see for this example that the final parameter estimates obtained are more accurate (continuous optimization) or no less accurate (stimulus-space winnowing) than full-grid search (Figure 10a through d).

Discussion

Summary of contributions

In recent years, there has been renewed interest in applying adaptive stimulus generation methods to study of sensory processing (Benda, Gollisch, Machens, & Herz, 2007; Paninski, Pillow, & Lewi, 2007; DiMattina & Zhang, 2013). However, with a few exceptions (Kujala & Lukka, 2006; Lesmes et al., 2010; Kim et al., 2014), adaptive stimulus generation methods which have been applied to psychophysical experiments have primarily focused on estimating or comparing low-dimensional models (Watson & Pelli, 1983; Kontsevich & Tyler, 1999; Prins, 2013b). In this methodological study, I present a detailed analysis of three implementations of the popular Ψ algorithm which generalize well to estimating multidimensional psychometric models. Two of these methods (Lookup- Ψ and Laplace- Ψ), to the best of my knowledge, represent novel approaches to psychophysical studies. Developing such efficient implementations for multidimensional experiments is particularly timely, as much recent effort has been focused on the problem of quantitatively describing how subjects combine multiple cues (for instance, combining cues to detect occlusion edges as in Figure 1) when making perceptual decisions (Landy & Kojima, 2001; Ernst & Banks, 2002; Körding & Wolpert, 2004; Ing et al., 2010; Saarela & Landy, 2012). When cues are combined in a nonlinear manner (Saunders & Knill, 2001; Frome et

al., 1981; Zhou & Mel, 2008), it becomes necessary to covary multiple stimulus parameters simultaneously in order to understand how they are integrated, and psychometric models for characterizing these interactions become more complex and harder to estimate.

Since the current study is methodological rather than intending to propose specific models of nonlinear cue combination (which will depend on the particular task being investigated), I demonstrate my implementations of the Ψ algorithm using a generic multivariate linear regression model (Equation 9) which readily generalizes the univariate model (Equation 2) commonly used in psychophysics (Kingdom & Prins, 2010). However, even with these generic models we readily observe the crucial importance of covarying parameters which interact in a nonadditive manner in order to accurately characterize their interaction (e.g., Figure 3), with numerous stimuli placed on the diagonals of the stimulus space for nonzero multiplicative interaction terms. In contrast, I found that for purely linear cue-combination rules, the Ψ method places stimuli entirely along the cardinal axes of the stimulus space (Supplementary Figure S3). This finding validates the standard procedure of characterizing perceptual sensitivity to each parameter dimension individually in cases of linear cue combination (Hillis et al., 2004).

Overview of methods

Broadly speaking, one may divide these procedures into methods that are based on a particle-filter representation of the posterior (Grid- Ψ , Prior- Ψ , and Lookup- Ψ) and a Laplace approximation to the posterior (Laplace- Ψ). In general, for models defined in lower dimensional parameter spaces (4–10 parameter dimensions), the particle-filter methods (Prior- Ψ and Lookup- Ψ) are perfectly suitable, and indeed may be preferable for models whose posterior density is poorly approximated by a Gaussian. The Prior- Ψ method provides perhaps the most straightforward approach to making the Ψ algorithm tractable in higher dimensions, and in contrast to the other methods I analyze here, it has been previously implemented in psychophysical studies (Kujala & Lukka, 2006). However, there are two potentially serious limitations to the particle-filter methods which merit discussion. The first limitation is that as the number of particles N_θ used to represent the posterior density grows, the time required by the Prior- Ψ procedure to generate the next stimulus can become experimentally inconvenient.

The second general limitation of particle-filter methods is that as the experiment progresses, the number of particles having probability significantly greater than 0 declines rapidly (Supplementary Figure S4; see also Bengtsson et al., 2008; Bickel et al., 2008;

Snyder et al., 2008). This can be problematic because an inaccurate representation of the posterior may make the integration over the posterior (Equation 15) required by the Ψ method unreliable. This problem can be rectified by periodic resampling of the posterior using MCMC methods (Gilks, 2005), but doing so can add time to the experiment. In the examples analyzed here, resampling was not necessary to obtain accurate parameter estimates, but a systematic investigation of MCMC resampling in the context of the methods analyzed in this paper is certainly an interesting avenue for future research. A strategy which has been used in previous work replaces MCMC simulation from a prior density with representation of the posterior density on adaptive sparse grids (Kim et al., 2014). This method has been shown to scale well with dimensionality in studies of econometric models (Heiss & Winschel, 2008; Winschel & Krätzig, 2010) and may provide a better approach to adaptively updating the particle filter than standard simulation-based methods.

I present a novel lookup-table approach making use of precomputed optimal stimuli and demonstrate that this method potentially offers a tremendous speedup over other particle-filter approaches to OED (i.e., Prior- Ψ), with no detrimental effects on estimation accuracy in my examples. Like all particle-based methods, it suffers from potential degeneration of the particle filter to a small number of effective particles as the experiment progresses (Bengtsson et al., 2008; Bickel et al., 2008; Snyder et al., 2008). However, assuming that the set of permissible designs $D = \{d_1, \dots, d_m\}$ remains fixed throughout the experiment, it should not be too time consuming upon resampling a new set of particles (much smaller the initial sample from the prior) to recompute the utility $U(d|\theta)$ of each design in D for each particle, particularly if the design set is winnowed to exclude designs of low expected utility as the experiment progresses. Furthermore, given a particle θ_i , for many models it may be possible to analytically compute an approximately optimal design (see Appendix). The Lookup- Ψ method is $O(N_d \cdot N_\theta)$ and demonstrates a substantial improvement over the $O(N_x \cdot N_\theta)$ Prior- Ψ method for an identical number N_θ of supports (for $N_d \ll N_\theta$), and I suggest that this method may be a powerful substitute for Prior- Ψ in situations where a large number of particles is needed to accurately represent the posterior density, thereby making Prior- Ψ intractable. Further exploration of this approach is potentially of great interest to researchers in statistical learning as well as in cognitive sciences.

Given the well-known limitations of particle filters in high-dimensional spaces (Snyder et al., 2008), implementations of OED for high-dimensional psychometric models may be best suited for a parametric representation of the posterior density by analytical forms. Indeed, this has been the approach taken by investi-

gators in computational neuroscience, who have used either Laplacian approximation methods (Lewi et al., 2009) or sum-of-Gaussians representations (DiMattina & Zhang, 2011) to represent evolving posterior densities. Making use of the log-concavity properties of the likelihood for many of the sigmoidal forms used in psychophysical research (Pratt, 1981), I demonstrate the potential utility of the Laplace approximation for experiments on nonlinear cue combination, which can give rise to models like Equation 9 with many more parameter dimensions than models typically used in psychophysical experiments (Figure 10; Kingdom & Prins, 2010). I discuss potentially fruitful applications of the Laplace- Ψ method to other kinds of perception and neuroscience experiments later.

Limitations of the present study

The fundamental limitation of this and any methodological study is that these algorithms are illustrated on a particular choice of models with particular assumptions about the stimuli, and therefore the benefits attained may be to some extent example dependent. Therefore, we should take the results presented here as providing an existence proof that such methods may be of benefit in some experiments, without claiming that they will necessarily be of benefit in all cases. Furthermore, for low-dimensional experiments (i.e., 1-D stimulus space), the standard implementation of Grid- Ψ (available in a MATLAB implementation at <http://www.palamedestoolbox.org/>) can easily choose stimuli quickly enough to be absorbed into a reasonably short interstimulus interval.

In this study, I presented three methods based on a discrete approximation of the posterior and one method based on a continuous approximation. Direct comparison between methods is difficult, since the IID case differs across each method. One comparison which I did make was to show that a relatively small number of particles sampled from a uniform prior (Prior- Ψ) allowed reasonably fast computation with little difference in final estimation error. I also demonstrated that for the same number of particles sampled from an identical prior, Lookup- Ψ was faster than Prior- Ψ without sacrifice in estimation accuracy. It was not the main goal of this paper to compare methods to each other but simply to define the methods, provide code for their implementation, and demonstrate that each potentially offers substantial improvement over the standard implementation of the Grid- Ψ method. In general, I feel that the Laplace- Ψ method is the best bet for generalization to higher dimensions, and preliminary results with simulated psychophysical estimation problems in even higher numbers of dimensions (tens of dimensions) have been promising. However, even the

Laplace- Ψ method runs into computational limitations as the number of observations increases, since evaluations of the likelihood become more costly as the experiment progresses.

Another limitation of the current work is that I did not consider estimation of the lapse rate. Previous studies have demonstrated that it may be important to estimate the lapse rate in order to get an accurate estimate of other parameters (Wichmann & Hill, 2001; Prins, 2012). In this study, I did not consider this problem, although it would be fairly straightforward to augment the parameter space and include the lapse rate.

Finally, a very important problem pertinent to any Bayesian approach that was not considered in this work is the issue of prior specification. One recent suggestion which has been shown to be highly effective in simulated psychophysical studies is to do hierarchical Bayesian modeling, where data from previous subjects is used to specify the prior for subsequent subjects (Kim et al., 2014). In this approach, parameters for the prior are specified by a hyper-prior, transforming the problem into that of specifying this hyper-prior. This innovative data-driven approach suggests a principled way to specify priors, and future experimental work along these lines to validate this approach should be pursued.

Future directions

As computational approaches become increasingly important in psychophysics and cognitive science, there are several important extensions of this work worthy of future investigation. In the present study, I considered the problem of how one can extend the popular Ψ method in order to efficiently estimate the parameters of higher dimensional psychometric models. The motivation for doing so is to develop methods which will permit us to estimate models of nonlinear visual cue combination (Figure 1). However, the appropriate nonlinear model to use is often unknown prior to experimentation, necessitating the fitting and comparison of multiple models (Pitt & Myung, 2002; Pitt, Myung, & Zhang, 2002; Myung & Pitt, 2009). Recent studies have developed adaptive stimulus optimization techniques for the goal of model comparison (Wang & Simoncelli, 2008; Cavagnaro et al., 2010; DiMattina & Zhang, 2011), and these have been applied experimentally in both human (Wang & Simoncelli, 2008; Cavagnaro et al., 2011) and animal (Tam, 2012) experiments. A very interesting direction for future research is to develop algorithms which combine the experimental goals of model estimation and comparison. One approach suggested by DiMattina and Zhang (2011) and implemented by Tam (2012) is to simply run

experiments in two phases: an estimation phase (E-phase), where stimuli optimized for each model are presented alternately, and a comparison phase (C-phase), where stimuli optimized for comparing models are presented. However, this is certainly not the only possibility, and it may be desirable to present stimuli which are optimized simultaneously for multiple experimental goals (Sugiyama & Rubens, 2008). This idea remains an open avenue for future research.

Most psychophysical and neurophysiological studies make use of parametric stimuli defined in low-dimensional spaces, for instance a bar of light or sinusoidal grating with a given orientation and spatial frequency (De Valois & De Valois, 1988). The responses elicited by these stimuli are used to estimate simple models having only a few parameters, like tuning curves or psychometric functions. However, an alternative approach to studying sensory systems which has gained a lot of traction in the neuroscience literature is the system-identification approach (Wu et al., 2006), where high-dimensional stimuli (up to hundreds of dimensions) are defined in a space corresponding to the activities of peripheral receptors for the modality in question (i.e., pixel space in vision). Similarly motivated experiments have been performed in the psychophysical literature in order to determine the perceptual filters that subjects use to detect a target or determine the position of a target in a noisy background (Ahumada, 1996; Mineault et al., 2009; Murray, 2011). It is of great interest for future work to see if adaptive stimulus generation methods, in particular the Laplace- Ψ method, can be fruitfully extended to high-dimensional psychophysical system-identification studies analogous to the application of such methods to neurophysiology (Lewi et al., 2009; DiMattina & Zhang, 2011). Along these same lines, another question is whether one can effectively use these methods to identify models of how neural populations are decoded by observers to make perceptual decisions. For instance, given a simulated (or simultaneously recorded) population of dozens or even hundreds of orientation-tuned neurons, can we learn a set of linear decoding weights which accurately predict subject performance in an orientation discrimination task (Berens et al., 2012)? This question is of great interest for future research.

Keywords: *adaptive stimulus generation, psychophysical methods, computational modeling*

Acknowledgments

The author would like to thank Kechen Zhang for his mentorship and many fruitful discussions over the years, and Giovanni Parmigiani for input on earlier

versions of the lookup-table idea. MATLAB/Octave code for these methods will be made available pending publication at <http://itech.fgcu.edu/cdimattina/>.

Commercial relationships: none.

Corresponding author: Christopher DiMattina.

Email: cdimattina@fgcu.edu.

Address: Computational Perception Laboratory, Department of Psychology, Florida Gulf Coast University, Fort Myers, FL, USA.

References

- Ahumada, A., Jr. (1996). Perceptual classification images from vernier acuity masked by noise. *Perception*, 26(Suppl 1), 18.
- Alspach, D. L., & Sorenson, H. W. (1972). Nonlinear Bayesian estimation using Gaussian sum approximations. *IEEE Transactions on Automatic Control*, 17(4), 439–448.
- Arulampalam, M. S., Maskell, S., Gordon, N., & Clapp, T. (2002). A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*, 50(2), 174–188.
- Atkinson, A. C., Donev, A. N., & Tobias, R. D. (2007). *Optimum experimental designs, with SAS, Vol. 34*. Oxford, UK: Oxford University Press.
- Bellman, R. E. (1961). *Adaptive control processes: A guided tour, Vol. 4*. Princeton, NJ: Princeton University Press.
- Benda, J., Gollisch, T., Machens, C. K., & Herz, A. V. (2007). From response to stimulus: Adaptive sampling in sensory physiology. *Current Opinion in Neurobiology*, 17(4), 430–436.
- Bengtsson, T., Bickel, P., & Li, B. (2008). Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems. In D. A. Nolan (Ed.), *Probability and statistics: Essays in honor of David A. Freedman* (pp. 316–334). Shaker Heights, OH: Institute of Mathematical Statistics.
- Berens, P., Ecker, A. S., Cotton, R. J., Ma, W. J., Bethge, M., & Tolias, A. S. (2012). A fast and simple population code for orientation in primate v1. *The Journal of Neuroscience*, 32(31), 10618–10626.
- Bickel, P., Li, B., & Bengtsson, T. (2008). Sharp failure rates for the bootstrap particle filter in high dimensions. In B. Salem Clarke & S. Ghosal (Eds.), *Pushing the limits of contemporary statistics: Contributions in honor of Jayanta K. Ghosh* (pp. 318–

- 329). Shaker Heights, OH: Institute of Mathematical Statistics.
- Bishop, C. M. (2006). *Pattern recognition and machine learning*. New York: Springer.
- Carpenter, J., Clifford, P., & Fearnhead, P. (1999). Improved particle filter for nonlinear problems. *IEE Proceedings—Radar, Sonar and Navigation*, 146(1), 2–7.
- Cavagnaro, D. R., Myung, J. I., Pitt, M. A., & Kujala, J. V. (2010). Adaptive design optimization: A mutual information-based approach to model discrimination in cognitive science. *Neural Computation*, 22(4), 887–905.
- Cavagnaro, D. R., Pitt, M. A., & Myung, J. I. (2011). Model discrimination through adaptive experimentation. *Psychonomic Bulletin & Review*, 18(1), 204–210.
- Chaloner, K., & Verdinelli, I. (1995). Bayesian experimental design: A review. *Statistical Science*, 10, 273–304.
- Cover, T. M., & Thomas, J. A. (2012). *Elements of information theory*. Hoboken, NJ: John Wiley & Sons.
- De Valois, R., & De Valois, K. (1988). Spatial vision. *Oxford: Oxford University Press*, 10, 145–175.
- DiMattina, C., Fox, S. A., & Lewicki, M. S. (2012). Detecting natural occlusion boundaries using local cues. *Journal of Vision*, 12(13):15, 1–21, doi:10.1167/12.13.15. [PubMed] [Article]
- DiMattina, C., & Zhang, K. (2011). Active data collection for efficient estimation and comparison of nonlinear neural models. *Neural Computation*, 23(9), 2242–2288.
- DiMattina, C., & Zhang, K. (2013). Adaptive stimulus optimization for sensory systems neuroscience. *Frontiers in Neural Circuits*, 7, 1–16.
- Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415(6870), 429–433.
- Frome, F. S., Buck, S. L., & Boynton, R. M. (1981). Visibility of borders: Separate and combined effects of color differences, luminance contrast, and luminance level. *Journal of the Optical Society of America*, 71(2), 145–150.
- Gilks, W. R. (2005). *Markov chain Monte Carlo*. Hoboken, NJ: Wiley Online Library.
- Hall, J. L. (1981). Hybrid adaptive procedure for estimation of psychometric functions. *The Journal of the Acoustical Society of America*, 69(6), 1763–1769.
- Heiss, F., & Winschel, V. (2008). Likelihood approximation by numerical integration on sparse grids. *Journal of Econometrics*, 144(1), 62–80.
- Hillis, J. M., Watt, S. J., Landy, M. S., & Banks, M. S. (2004). Slant from texture and disparity cues: Optimal cue combination. *Journal of Vision*, 4(12):1, 967–992, doi:10.1167/4.12.1. [PubMed] [Article]
- Ing, A. D., Wilson, J. A., & Geisler, W. S. (2010). Region grouping in natural foliage scenes: Image statistics and human performance. *Journal of Vision*, 10(4):10, 1–19, doi:10.1167/10.4.10. [PubMed] [Article]
- Kay, S. M. (1993). *Fundamentals of statistical signal processing*. Englewood Cliffs, NJ: PTR Prentice-Hall.
- Kim, W., Pitt, M. A., Lu, Z.-L., Steyvers, M., & Myung, J. I. (2014). A hierarchical adaptive approach to optimal experimental design. *Neural Computation*, 26(11), 2465–2492.
- Kingdom, F., & Prins, N. (2010). *Psychophysics: A practical introduction*. London: Academic Press.
- Knill, D. C., & Pouget, A. (2004). The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends in Neurosciences*, 27(12), 712–719.
- Knill, D. C., & Saunders, J. A. (2003). Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vision Research*, 43(24), 2539–2558.
- Konishi, S., Yuille, A. L., Coughlan, J. M., & Zhu, S. C. (2003). Statistical edge detection: Learning and evaluating edge cues. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(1), 57–74.
- Kontsevich, L. L., & Tyler, C. W. (1999). Bayesian adaptive estimation of psychometric slope and threshold. *Vision Research*, 39(16), 2729–2737.
- Körding, K. P., & Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature*, 427(6971), 244–247.
- Kujala, J. V., & Lukka, T. J. (2006). Bayesian adaptive estimation: The next dimension. *Journal of Mathematical Psychology*, 50(4), 369–389.
- Kuss, M., Jakel, F., & Wichmann, F. A. (2005). Bayesian inference for psychometric functions. *Journal of Vision*, 5(5):8, 478–492, doi:10.1167/5.5.8. [PubMed] [Article]
- Landy, M. S., & Kojima, H. (2001). Ideal cue combination for localizing texture-defined edges. *Journal of the Optical Society of America A*, 18(9), 2307–2320.
- Lesmes, L. A., Lu, Z.-L., Baek, J., & Albright, T. D. (2010). Bayesian adaptive estimation of the contrast sensitivity function: The quick CSF method.

- Journal of Vision*, 10(3):17, 1–21, doi:10.1167/10.3.17. [PubMed] [Article]
- Lewi, J., Butera, R., & Paninski, L. (2009). Sequential optimal design of neurophysiology experiments. *Neural Computation*, 21(3), 619–687.
- Lu, Z.-L., & Doshier, B. (2013). *Visual psychophysics: From laboratory to theory*. Cambridge, MA: MIT Press.
- Maloney, L. T. (1990). Confidence intervals for the parameters of psychometric functions. *Perception & Psychophysics*, 47(2), 127–134.
- Martin, D. R., Fowlkes, C. C., & Malik, J. (2004). Learning to detect natural image boundaries using local brightness, color, and texture cues. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(5), 530–549.
- McGraw, P. V., Whitaker, D., Badcock, D. R., & Skillen, J. (2003). Neither here nor there: Localizing conflicting visual attributes. *Journal of Vision*, 3(4):2, 265–273, doi:10.1167/3.4.2. [PubMed] [Article]
- Mineault, P. J., Barthelme, S., & Pack, C. C. (2009). Improved classification images with sparse priors in a smooth basis. *Journal of Vision*, 9(10):17, 1–24, doi:10.1167/9.10.17. [PubMed] [Article]
- Murphy, K. P. (2012). *Machine learning: A probabilistic perspective*. Cambridge, MA: MIT Press.
- Murray, R. F. (2011). Classification images: A review. *Journal of Vision*, 11(5):2, 1–25, doi:10.1167/11.5.2. [PubMed] [Article]
- Myung, J. I., & Pitt, M. A. (2009). Optimal experimental design for model discrimination. *Psychological Review*, 116(3), 499–518.
- O'Hagan, A., & Kingman, J. (1978). Curve fitting and optimal design for prediction. *Journal of the Royal Statistical Society: Series B (Methodological)*, 40(1), 1–42.
- Paninski, L., Pillow, J., & Lewi, J. (2007). Statistical models for neural encoding, decoding, and optimal stimulus design. *Progress in Brain Research*, 165, 493–507.
- Pitt, M. A., & Myung, I. J. (2002). When a good fit can be bad. *Trends in Cognitive Sciences*, 6(10), 421–425.
- Pitt, M. A., Myung, I. J., & Zhang, S. (2002). Toward a method of selecting among computational models of cognition. *Psychological Review*, 109(3), 472–491.
- Pratt, J. W. (1981). Concavity of the log likelihood. *Journal of the American Statistical Association*, 76(373), 103–106.
- Prins, N. (2012). The psychometric function: The lapse rate revisited. *Journal of Vision*, 12(6):25, 1–16, doi:10.1167/12.6.25. [PubMed] [Article]
- Prins, N., & Kingdom, F. A. A. (2009). Palamedes: Matlab routines for analyzing psychophysical data. Available from <http://www.palamedestoolbox.org>.
- Prins, N. (2013b). The psi-marginal adaptive method: How to give nuisance parameters the attention they deserve (no more, no less). *Journal of Vision*, 13(7):3, 1–17, doi:10.1167/13.7.3. [PubMed] [Article]
- Saarela, T. P., & Landy, M. S. (2012). Combination of texture and color cues in visual segmentation. *Vision Research*, 58, 59–67.
- Saunders, J. A., & Knill, D. C. (2001). Perception of 3D surface orientation from skew symmetry. *Vision Research*, 41(24), 3163–3183.
- Skottun, B. C., Bradley, A., Sclar, G., Ohzawa, I., & Freeman, R. D. (1987). The effects of contrast on visual orientation and spatial frequency discrimination: A comparison of single cells and behavior. *Journal of Neurophysiology*, 57(3), 773–786.
- Snyder, C., Bengtsson, T., Bickel, P., & Anderson, J. (2008). Obstacles to high-dimensional particle filtering. *Monthly Weather Review*, 136(12), 4629–4640.
- Sugiyama, M., & Rubens, N. (2008). A batch ensemble approach to active learning with model selection. *Neural Networks*, 21(9), 1278–1286.
- Tam, W. (2012). *Adaptive modeling of marmoset inferior colliculus neurons in vivo* (Unpublished doctoral dissertation). The Johns Hopkins University School of Medicine, Baltimore, MD.
- Trommershauser, J., Kording, K., and Landy, M. S. (2011). *Sensory cue integration*. Oxford, UK: Oxford University Press.
- Vogels, R., & Orban, G. (1990). How well do response changes of striate neurons signal differences in orientation: A study in the discriminating monkey. *The Journal of Neuroscience*, 10(11), 3543–3558.
- Wang, Z., & Simoncelli, E. P. (2008). Maximum differentiation (mad) competition: A methodology for comparing computational models of perceptual quantities. *Journal of Vision*, 8(12):8, 1–13, doi:10.1167/8.12.8. [PubMed] [Article]
- Watson, A. B., & Pelli, D. G. (1983). Quest: A Bayesian adaptive psychometric method. *Perception & Psychophysics*, 33(2), 113–120.
- Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics*, 63(8), 1293–1313.
- Winschel, V., & Krätzig, M. (2010). Solving, estimating, and selecting nonlinear dynamic models without the curse of dimensionality. *Econometrica*, 78(2), 803–821.
- Wu, M. C.-K., David, S. V., & Gallant, J. L. (2006). Complete functional characterization of sensory

neurons by system identification. *Annual Review of Neuroscience*, 29, 477–505.

Zavitz, E., & Baker, C. L. (2014). Higher order image structure enables boundary segmentation in the absence of luminance or contrast cues. *Journal of Vision*, 14(4):14, 1–14, doi:10.1167/14.4.14. [PubMed] [Article]

Zhou, C., & Mel, B. W. (2008). Cue combination and color edge detection in natural scenes. *Journal of Vision*, 8(4):4, 1–25, doi:10.1167/8.4.4. [PubMed] [Article]

Appendix

Here I present the derivation of the D-optimal design for estimating the univariate psychometric model (Equation 2), along with an analytical approximation to this design. The D-optimal design $d = \{x_1, x_2\}$ is the set of observations which maximize the determinant of the Fisher information matrix, given by

$$I_F(d, \theta) = i_F(x_1, \theta) + i_F(x_2, \theta), \quad (29)$$

where

$$i_F(x, \theta) = \left\langle -\frac{d^2}{d\theta^2} \ln p(r|x, \theta) \right\rangle_r. \quad (30)$$

Assuming two possible subject responses ($r = 0, 1$), we have $p(r=1|\mathbf{x}, \theta) = \sigma(\theta^T \mathbf{x}) = \eta$ and $p(r=0|\mathbf{x}, \theta) = 1 - \eta$, where we use the simplifying notation $\mathbf{x} = (1, x)^T$. Applying the definition in Equation 30, it is simple to show that

$$i_F(\mathbf{x}, \theta) = \sigma'(\mathbf{x}^T \theta) \begin{bmatrix} 1 & x \\ x & x^2 \end{bmatrix}, \quad (31)$$

and therefore

$$I_F(d, \theta) = \sigma'(\mathbf{x}_1^T \theta) \begin{bmatrix} 1 & x_1 \\ x_1 & x_1^2 \end{bmatrix} + \sigma'(\mathbf{x}_2^T \theta) \begin{bmatrix} 1 & x_2 \\ x_2 & x_2^2 \end{bmatrix}. \quad (32)$$

A little algebra shows that

$$\det(I_F(d, \theta)) = \alpha_1 \alpha_2 (x_1 - x_2)^2, \quad (33)$$

where $\alpha_1 = \sigma'(\mathbf{x}_1^T \theta)$ and $\alpha_2 = \sigma'(\mathbf{x}_2^T \theta)$. Equation 33 is the function plotted in Figure 6b for $\theta = (0, 1)^T$.

It is not possible to analytically solve Equation 33 for the optimal design d^* , due to the presence of transcendental functions, but using the (λ, β) parameterization (Equation 4) and making some approximations allows us to solve analytically for an approximately D-optimal design. To see this, assume that the solution is symmetric about the threshold λ , so that $x_1 = \lambda + v$ and $x_2 = \lambda - v$. Substituting into Equation 33 and using the parameterization (Equation 4) yields

$$\det I_F(v) = 4 \left(\sigma'(\beta v) \right)^2 v^2, \quad (34)$$

where we make use of fact that $\sigma'(-u) = \sigma'(u)$. Since $\det(I_F(v)) > 0$, we can instead optimize

$$H(v) = \sqrt{\det I_F(v)} = \frac{v}{1 + \cosh(\beta v)}, \quad (35)$$

making use of the identity

$$\sigma'(u) = \frac{1}{2} \cdot \frac{1}{1 + \cosh(u)}. \quad (36)$$

Differentiating Equation 35 and setting it equal to 0 yields

$$\frac{1}{\cosh(\beta v)} + 1 = \beta v \tanh(\beta v). \quad (37)$$

For small arguments, we have $\cosh(u) \approx 1$ and $\tanh(u) \approx u$, yielding the final result

$$\hat{v} = \frac{\sqrt{2}}{\beta}. \quad (38)$$

For the example shown in Figure 6a with $(\lambda, \beta) = (0, 1)$, this gives us an approximate optimal design of $\hat{d}^* = \{\pm 1.41\}$, which is reasonably close to the numerically computed D-optimal design $d^* = \{\pm 1.54\}$.