European Journal of Applied Mathematics (2025), 1–27
doi:10.1017/S0956792525000075

CAMBRIDGE UNIVERSITY PRESS

PAPER

Numerical solution of a PDE arising from prediction with expert advice

Jeff Calder[1], Nadejda Drenska[2] and Drisana Mosaphir[1]

[1]School of Mathematics, University of Minnesota, Minneapolis, MN, USA
[2]Department of Mathematics, Louisiana State University, Baton Rouge, LA, USA
Corresponding author: Jeff Calder; Email: jwcalder@umn.edu

Received: 09 June 2024; **Revised:** 02 February 2025; **Accepted:** 10 February 2025

Keywords: Online prediction; COMB strategy; viscosity solutions; nonlinear elliptic PDE; monotone finite difference scheme

2020 Mathematics Subject Classification: 35Q68, 65N12, 65N06 (Primary); 35D40, 35J60 (Secondary)

Abstract

This work investigates the online machine learning problem of prediction with expert advice in an adversarial setting through numerical analysis of, and experiments with, a related partial differential equation. The problem is a repeated two-person game involving decision-making at each step informed by n experts in an adversarial environment. The continuum limit of this game over a large number of steps is a degenerate elliptic equation whose solution encodes the optimal strategies for both players. We develop numerical methods for approximating the solution of this equation in relatively high dimensions ($n \leq 10$) by exploiting symmetries in the equation and the solution to drastically reduce the size of the computational domain. Based on our numerical results we make a number of conjectures about the optimality of various adversarial strategies, in particular about the non-optimality of the COMB strategy.

1. Introduction

This paper is focused on the classical online learning problem of prediction with expert advice. Given the advice of n experts who each make predictions in real time about an unknown time-varying quantity (e.g., the price of a stock or option at some time in the future), a player must decide which expert's advice to follow. The problem is often formulated in an *online* setting, whereby at each step of the game, the player has knowledge of the historical performance of each expert and may use this information to decide which expert to follow at that step. The overall goal is to perform as well as the best-performing expert, or as close to this as possible. We are particularly interested in the adversarial setting, where the performance of the experts is controlled by an adversary, whose goal is to minimize the gains of the player.

A simple and common way to formulate the game is to assume each expert's prediction is either correct or incorrect at each time step. Given n experts, this can be described by a binary vector $\mathbf{v} \in \mathcal{B}^n$, $\mathcal{B} := \{0, 1\}$, where $v_i = 1$ if expert i is correct, and $v_i = 0$ otherwise. The player gains 1 if the expert they chose made a correct prediction, and gains nothing otherwise; thus the player gains v_i if they follow expert i. The performance of the player is often measured with the notion of *regret* with respect to each expert, which is the difference between a given expert's gains and the player's gains. We let $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ denote the regret vector with respect to all n experts, so x_i is the regret to the i^{th} expert, or rather, the number of times the i^{th} expert was correct minus the number of times the player was correct. After T steps, the gain of the player and any given expert is at most T, and the worst-case regret with respect to any given expert is at most T. The goal of the player is to minimize their regret with respect to the best-performing expert; thus, the player would like to minimize $\max\{x_1, \ldots, x_n\}$ at the end of the game. The goal of mathematical and numerical analysis is to compute or approximate the optimal player

strategies (i.e., determine which expert the player should follow). There are two standard approaches for how long the game is played; the finite horizon setting, where the game is played for a fixed number of steps T, and the geometric horizon setting, where the game ends with probability $\delta > 0$ at each step. In the geometric horizon setting, the number of steps of the game is a random variable following the geometric distribution with parameter δ. In this work, we focus on the geometric horizon problem, but we expect our techniques to work in the finite horizon setting as well.

In order to proceed further, we need to make a modelling assumption on the experts. In this paper, we follow the convention of a *worst-case* analysis where we assume the experts are controlled by an adversary whose goal is to maximize the player's regret at the end of the game. The adversarial setting yields a two-player zero sum game and introduces another mathematical problem of determining the optimal strategies for the adversary. We also focus on the more general setting of *mixed strategies* where the player and adversary both employ randomized strategies. At each step, the player chooses a probability distribution α over the experts $\{1, \dots, n\}$, and the expert followed by the player is drawn independently (from other steps and the adversary's choices) from the distribution α. Likewise, the adversary chooses a probability distribution β over the binary sequences $\mathbf{v} \in \mathcal{B}^n$, and the experts are advanced by drawing a sample $\mathbf{v} \in \mathcal{B}^n$ from the distribution β. The problem of determining optimal strategies then boils down to deciding how the player and market should set their probability distributions α and β at each time step, given the current regret vector $x \in \mathbb{R}^n$. The goal of the player is to minimize their expected regret, while the adversary's goal is to maximize the expected regret.

It is a difficult problem to determine the optimal strategies for the player and the adversary for a general number of experts n and a large number of steps T in the game. Initial results were established in Cover's original paper [18] for $n = 2$ experts, and more recently in [27] for $n = 3$ experts. A breakthrough occurred in a series of papers by Drenska and Kohn [20, 22, 23], who took the perspective that the prediction from expert advice problem is a discrete analogue of two-player differential games [26]. They formulated a value function for the game and showed that as the number of steps T of the game tends to infinity, the rescaled value function converges to the *viscosity solution* $u \in C(\mathbb{R}^n)$ of the degenerate elliptic partial differential equation (PDE)

$$u(x) - \frac{1}{2} \max_{\mathbf{v} \in \mathcal{B}^n} \mathbf{v}^T \nabla^2 u(x)\, \mathbf{v} = \max\{x_1, \dots, x_n\} \quad \text{for} \quad x \in \mathbb{R}^n. \tag{1.1}$$

The state variable $x \in \mathbb{R}^n$ of the PDE (1.1) is the regret vector at the start of the game, and the solution of the PDE $u(x)$ is the worst-case regret over all the experts at the end of the game, provided each player plays optimally (and in the limit as $T \to \infty$). Thus, the long-time behaviour of the adversarial prediction with expert advice problem, and the corresponding asymptotically optimal strategies for the player and adversary, can be determined by solving a PDE! It turns out that the optimal player strategy is to choose the probability distribution $\alpha(x) = \nabla u(x)$, while the optimal adversarial strategy involves the binary vectors $\mathbf{v}$ saturating the maximum over $\mathbf{v} \in \mathcal{B}^n$ in (1.1) (we give more detail in Section 1.1).

In this paper, we develop numerical methods for approximating the viscosity solution of (1.1), so that we may shed light on the optimal strategies for the player and adversary. There are two challenging aspects of solving (1.1). First, the PDE is posed on all of $\mathbb{R}^n$ and does not have any kind of natural restriction to a compact computational domain with boundary conditions. To address this, we prove a localization result for (1.1) showing that the domain may be restricted to a box $\Omega_T = [-T, T]^n$ and that errors in the Dirichlet boundary condition $u|_{\partial\Omega_T}$ *do not* propagate far into the interior of Ω, allowing us to obtain an accurate solution sufficiently interior to Ω_T (precise results are given later). The second challenge is that the theory is fairly complete for the $n = 2, 3, 4$ expert problems, so the interesting cases to study are in the fairly high dimensional setting of $n \geq 5$ experts, where it is generally difficult to solve PDEs on regular grids or meshes, due to the curse of dimensionality. To overcome this issue, we exploit symmetries in the equation (1.1)–in particular, permutation invariance of the coordinates $x_1, \dots, x_n$ – in order to restrict the computational domain to the sector in $\mathbb{R}^{n-1}$ where $T \geq x_1 \geq x_2 \geq \cdots x_{n-1} \geq x_n = 0$. We develop a numerical method that allows us to restrict all computations to this sector, which has vanishingly small measure compared to the whole box Ω_T. This allows us to numerically solve the

https://doi.org/10.1017/S0956792525000075 Published online by Cambridge University Press

PDE (1.1) on reasonably fine grids up to dimension $n = 10$. Based on our numerical results, we make a number of conjectures about optimality of various strategies. We summarize these results in Section 1.2 after giving a more thorough description of the background material.

1.1 Background

Online prediction with expert advice is an example of a sequential decision-making problem and has applications in algorithm boosting [25], stock price prediction and portfolio optimization [25], self-driving car software [2], and many other problems. The prediction with expert advice problem originated in the works of Cover [18] and Hannan [29], who provided optimal strategies for the two expert problems. In the intervening years, much attention has been focused on heuristic algorithms that give good performance, but may not be optimal. A commonly used algorithm is the *multiplicative weights algorithm*, in which the player maintains a set of positive weights $w_1, \dots, w_n$ for each expert that are used to form a weighted average of the expert advice, and the weights are updated in real time based on the performance of each expert. For the finite horizon problem, it has been shown [16] that the multiplicative weights algorithm is optimal in an asymptotic sense, as the number of experts n and number of steps of the game T *both* tend to infinity, but fails to be optimal when the numbers of experts are fixed and finite, as is the case in practice. Optimal algorithms for the geometric horizon and finite horizon problems for $n \leq 3$ experts were developed and studied in Gravin et al. [27] and Abbasi et al. [1]. Additionally, lower bounds for regrets in a broader class than multiplicative weights have been shown [28]. Further work on algorithms for both the finite and geometric horizon problems is contained in [1, 15, 16, 30, 36, 42].

Recently, attention has shifted back to the problem of optimal strategies for a finite number of experts [5–8, 12, 20–23, 27]. The focus of this paper is on the problem with *mixed strategies* (i.e., random strategies) against an adversarial environment that was briefly described in the previous section. We now describe this setting in more detail. We have n experts making predictions, a player who chooses at each step which expert to follow, and an adversarial environment that decides which experts gain or lose at each step of the game. The strategies are *mixed*, so both the player and the adversary choose probability distributions over their possible plays, and their actual plays are random variables drawn from those distributions. At the k^{th} step of the game, the player chooses a probability distribution α_k over the experts $\mathcal{A}_n := \{1, 2, \dots, n\}$ and the adversary chooses a probability distribution β_k over the binary sequences $\mathcal{B}^n = \{0, 1\}^n$. Random variables from these distributions $i_k \sim \alpha_k$ and $\mathbf{v}_k \sim \beta_k$ are drawn, and the player chooses to follow expert i_k and the market advances the experts corresponding to the positions of the ones in the binary vector $\mathbf{v}_k$.

The performance of the player is measured by their regret to each expert, which is the difference between the expert's gains and those of the player. We let $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ denote the regret vector, so x_i is the current regret with respect to expert i. Let us write the coordinates of $\mathbf{v}_k$ as $\mathbf{v}_k = (v_{k,1}, v_{k,2}, \dots, v_{k,n})$. Then on the k^{th} step the player accumulates regret of $v_{k,j} - v_{k,i_k}$ with respect to expert j. If the regret vector started at $x \in \mathbb{R}^n$ on the first step of the game, then the regret after T steps is

$$R_T := x + \sum_{k=1}^{T} (\mathbf{v}_k - v_{k,i_k} \mathbb{1}).$$

At the end of the game, the regret R_T is evaluated with a payoff function $g{:}\mathbb{R}^n \to \mathbb{R}$. The player's goal is to minimize $g(R_T)$ while the adversary's goal is to maximize $g(R_T)$. The most commonly used payoff is the maximum regret $g(x) = \max\{x_1, x_2, \dots, x_n\}$, which simply reports the regret of the player with respect to the best-performing expert. The game can be played in the finite horizon setting where T is fixed, or the geometric horizon setting where the game ends with probability $\delta > 0$ at each step, so T is a random variable.

We focus on the geometric horizon problem. In this case, the value function $U_\delta{:}\mathbb{R}^n \to \mathbb{R}$ is defined by

$$U_\delta(x) = \inf_{\alpha} \sup_{\beta} \mathbb{E}\left[g\left(x + \sum_{k=1}^{T} (\mathbf{v}_k - v_{k,i_k} \mathbb{1}) \right) \right], \tag{1.2}$$

https://doi.org/10.1017/S0956792525000075 Published online by Cambridge University Press

where $\alpha = (\alpha_1, \alpha_2, \dots)$, $\beta = (\beta_1, \beta_2, \dots)$, and T is the time at which the game stops, which is a geometric random variable with parameter δ. The value function $U_\delta(x)$ is the expected value of the payoff at the end of the game given that the regret vector starts at $x \in \mathbb{R}^n$ on the first step and both players play optimally. The inf and the sup are over *strategies for the players*, which enforce that α_k and β_k depend only on the current value of the regret and the past choices of both players. The value function is unchanged by swapping the inf and sup in (1.2).

It was shown in [20, 22] that the rescaled value functions

$$u_\delta(x) := \sqrt{\delta}\, U_\delta\left(\frac{x}{\delta}\right)$$

converge locally uniformly, as $\delta \to 0$, to the viscosity solution of the degenerate elliptic PDE

$$u - \frac{1}{2}\max_{\mathbf{v}\in B_n} \mathbf{v}^T \nabla^2 u\, \mathbf{v} = g \quad \text{on } \mathbb{R}^n, \tag{1.3}$$

which is the same as (1.1) except with a general payoff g. The PDE (1.3) contains all of the information about the prediction problem in the asymptotic regime where the number of steps T tends to ∞, which is equivalent to sending the geometric stopping probability $\delta \to 0$. It was shown in [22] that the asymptotically optimal player strategy is to use the probability distribution[1]

$$\alpha_k = \nabla u(x), \tag{1.4}$$

where $x \in \mathbb{R}^n$ is the current regret on the k^{th} step of the game, and the optimal adversary strategy is to choose

$$\mathbf{v}_k \in \operatorname{argmax}_{\mathbf{v}\in B^n}\{\mathbf{v}^T \nabla^2 u(x)\mathbf{v}\}, \tag{1.5}$$

and then advance the experts in $\mathbf{v}_k$ with probability $\frac{1}{2}$ and those in $\mathbb{1} - \mathbf{v}_k$ with probability $\frac{1}{2}$. Notice that the adversarial strategy $\mathbf{v}$ is equivalent to $\mathbb{1} - \mathbf{v}$.

Determining the adversary's optimal strategies in the max-regret setting, where we take $g(x) = \max\{x_1, \dots, x_n\}$, is an open problem for a general number of experts n. It was conjectured in [27] that the COMB strategy of ordering the experts by regret and using the alternating zeros and ones vector $\mathbf{v} = (0, 1, 0, 1, \dots)$, which resembles the teeth of a comb, is asymptotically optimal as $T \to \infty$ for all n, though this was proven in [27] only for $n = 2$ and $n = 3$ experts. The idea behind the COMB strategy is to group the experts into as equally matched groups as possible and advance one group or the other with equal probability, which makes it difficult for the player to gain any advantage. However, since the experts are ordered by regret $x_1 \geq x_2 \geq \cdots \geq x_n$, they are essentially ordered by performance, and so the even-numbered experts are slightly worse on average compared to the odd-numbered experts. Hence, there is reason to believe that COMB can be improved upon for larger numbers of experts by modifying the COMB vector slightly.

The PDE perspective developed in [20] can help shed light on the optimal adversarial strategy, since it turns out we can derive *explicit* solutions for the PDE (1.1) for $n \leq 4$ experts. Below, we present the solutions in the sector

$$\mathbb{S}_n = \{x \in \mathbb{R}^n : x_1 \geq x_2 \geq \cdots \geq x_n\}.$$

Since the PDE is unchanged under permutations of the coordinates, this completely determines the solution. It was shown in [20] that for $n = 2$ experts the solution of (1.1) is given in the sector $\mathbb{S}_2$ by

$$u(x) = x_1 + \frac{1}{2\sqrt{2}} e^{\sqrt{2}(x_2 - x_1)}, \tag{1.6}$$

and for $n = 3$, the solution is given in $\mathbb{S}_3$ by

$$u(x) = x_1 + \frac{1}{2\sqrt{2}} e^{\sqrt{2}(x_2 - x_1)} + \frac{1}{6\sqrt{2}} e^{\sqrt{2}(2x_3 - x_2 - x_1)}. \tag{1.7}$$

[1] As we will see later in the paper, the solution u of (1.3) satisfies $\nabla u \cdot \mathbb{1} = 1$, and is monotonically increasing, so α_k is a probability vector.

https://doi.org/10.1017/S0956792525000075 Published online by Cambridge University Press

Since we have explicit solutions, we can check the optimal adversarial strategies via (1.5) within the sector $\mathbb{S}_n$. For $n=2$ both $\mathbf{v}=(1,0)$ and $\mathbf{w}=(0,1)$ are globally optimal for all $x\in\mathbb{S}_n$, which are both the same COMB strategy since $\mathbf{v}=\mathbb{1}-\mathbf{w}$. For $n=3$, both $(1,0,0)$ and $(0,1,0)$ are globally optimal (as well as a their equivalent strategies $(0,1,1)$ and $(1,0,1)$, which we will omit from now on). The second strategy $(0,1,0)$ is the COMB strategy, while the first $(1,0,0)$ is not.

For $n=4$ experts, it was shown in [7] that the solution of (1.1) is given by

$$\begin{aligned} u(x) = x_1 &- \frac{\sqrt{2}}{4}\sinh(\sqrt{2}(x_1-x_2)) \\ &+ \frac{\sqrt{2}}{2}\arctan\left(e^{\frac{x_4+x_3-x_2-x_1}{\sqrt{2}}}\right)\cdot \\ &\cosh\left(\frac{x_4-x_3+x_2-x_1}{\sqrt{2}}\right)\cosh\left(\frac{-x_4+x_3+x_2-x_1}{\sqrt{2}}\right)\cosh\left(\frac{-x_4-x_3+x_2+x_1}{\sqrt{2}}\right) \\ &+ \frac{\sqrt{2}}{2}\operatorname{arctanh}\left(e^{\frac{x_4+x_3-x_2-x_1}{\sqrt{2}}}\right)\cdot \\ &\sinh\left(\frac{x_4-x_3+x_2-x_1}{\sqrt{2}}\right)\sinh\left(\frac{-x_4+x_3+x_2-x_1}{\sqrt{2}}\right)\sinh\left(\frac{-x_4-x_3+x_2+x_1}{\sqrt{2}}\right), \end{aligned} \tag{1.8}$$

from which it is possible to prove [7] that the COMB strategy $(1,0,1,0)$ is optimal, as well as the non-COMB strategy $(0,1,1,0)$. For $n\geq 5$ experts, an explicit solution is unknown, and the question of the optimality of the COMB strategy is open.

It is worthwhile mentioning that it is remarkable that explicit solutions have been obtained for the degenerate elliptic PDE (1.1) for $n\leq 4$ dimension. Usually, explicit solutions for nonlinear PDE are not available. Furthermore, the solutions given in (1.6), (1.7) and (1.8) are all twice continuously differentiable with Lipschitz second partial derivatives, i.e., they are classical $C^{2,1}(\mathbb{R}^n)$ solutions. This is also remarkable, since one would only expect this for *uniformly elliptic* equations (since the right-hand side is Lipschitz continuous) and not for degenerate elliptic equations. As far as we are aware, there is no general regularity theory that can explain this.

In fact, the existence of classical solutions of (1.1) is closely tied to the existence of a globally optimal strategy. As a simple example, for $n=2$ experts the strategy $\mathbf{v}=(1,0)$ is optimal in $\mathbb{S}_2$, and so the PDE (1.1) reduces to the one-dimensional linear equation

$$u-\frac{1}{2}u_{x_1x_1}=x_1 \quad \text{in } \mathbb{S}_2.$$

We can integrate this, keeping only the exponentially decaying solution, to obtain

$$u(x)=x_1+f(x_2)e^{-\sqrt{2}x_1} \quad \text{on } \mathbb{S}_2.$$

Using the symmetry $u(x_1,x_2)=u(x_2,x_1)$ we obtain $u_{x_1}=u_{x_2}$ on $\partial\mathbb{S}_2=\{x\in\mathbb{R}^2 : x_1=x_2\}$. This yields

$$1-\sqrt{2}f(x)e^{-\sqrt{2}x}=u_{x_1}(x,x)=u_{x_2}(x,x)=f'(x)e^{-\sqrt{2}x},$$

and so

$$f(x)=Ce^{-\sqrt{2}x}+\frac{1}{2\sqrt{2}}e^{\sqrt{2}x}.$$

Since $u_{x_1}(0)=u_{x_2}(0)=\frac{1}{2}$ we have $f(0)=\frac{1}{2\sqrt{2}}$ and so $C=0$. Substituting this above yields the two expert solutions u given by (1.6). Roughly the same procedure can be carried out for $n=3$ and $n=4$ experts, though the $n=4$ case is particularly tedious (see [7]).

The question of whether the COMB strategy is asymptotically optimal for $n\geq 5$ experts is an open problem. Some recent work [17] gives experimental numerical evidence that COMB is *not optimal* for $n=5$ experts. The numerical experiments in [17] simulated the two-player game and compared the COMB strategy against the strategy $(1,0,1,0,0)$, the latter appearing to be strictly better. It was shown in [31, 32] that COMB is at least as powerful as the setting of randomly choosing which half of the

https://doi.org/10.1017/S0956792525000075 Published online by Cambridge University Press

experts to advance (the so-called Bernoulli strategy). This motivates our work of numerically solving the PDE (1.1) in order to shed light on the optimality of COMB and other strategies for $n \geq 5$ experts.

We mention that an analogous parabolic PDE exists for the finite-horizon setting of this problem [20], motivating a similar treatment in terms of numerics for the parabolic equation as well. Some progress in the parabolic case can be found in [5]. Further, related settings such as prediction against a limited adversary [6] (rather than the optimal adversary being studied in this work) and malicious experts [8] have been studied as well. A related PDE where $\mathbf{v}$ is chosen just from the standard basis vectors $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ has been studied and has a closed-form solution for a general number n of experts [31, 32]. Additionally, some PDE approaches to online learning problems and neural network-based problems can be found in [43]. We also mention that repeated two-player games also appear in the PDE literature in multiple other settings [3, 4, 9, 11, 13, 14, 24, 33–35, 37, 39, 40].

1.2 Main results and conjectures

In Section 4, we present the results of numerical experiments solving the prediction with expert advice PDE (1.1) for $n \leq 10$ experts. We summarize the main results we obtain from the numerical experiments here.

1. We have strong numerical evidence that the COMB strategy is not globally optimal for $5 \leq n \leq 10$. This validates the numerical evidence from [17] for $n = 5$ experts.
2. For $n = 5$ experts, we have strong numerical evidence that the non-COMB adversarial strategy $(0, 1, 0, 1, 1)$, equivalent to $(1, 0, 1, 0, 0)$, is the only globally optimal strategy. This is the same strategy that was numerically shown to be better than COMB in [17].
3. For $6 \leq n \leq 10$ experts, we have strong numerical evidence that there are no globally optimal adversary strategies.

From these numerical results, we state a number of conjectures that we leave for future work.

Conjecture 1.1. *The COMB strategy is globally asymptotically optimal (i.e., on the sector $\mathbb{S}_n$) only for $n \leq 4$ experts.*

Conjecture 1.2. *The non-COMB strategy $(1, 0, 1, 0, 0)$ is the only globally asymptotically optimal strategy for $n = 5$ experts.*

Conjecture 1.3. *There is no globally asymptotically optimal adversary strategy that is constant on the sector $\mathbb{S}_n$ for $n \geq 6$ experts.*

Recalling our discussion in Section 1.1 about the connection between explicit solutions of the PDE (1.1) and the existence of optimal strategies, if Conjecture 1.2 is true, then we expect there to exist a classical explicit solution of the $n = 5$ expert PDE, similar to the solutions given in (1.6), (1.7), and (1.8) for the $n = 2, 3, 4$ expert problems.

1.2.1 Open problem

Determine an analytic expression for the solution of the PDE (1.1) for $n = 5$.

An explicit solution for $n = 5$ experts would allow one to check the validity of Conjecture 1.2, as was done for $n = 4$ experts in [7]. If Conjecture 1.3 is true, then this strongly suggests that it will be impossible to find an explicit solution of the PDE for $n \geq 6$ experts and that the solution may fail to be a classical solution in $C^{2,1}(\mathbb{R}^n)$ when $n \geq 6$. If there is no globally optimal strategy, then there will be regions of $\mathbb{S}_n$ corresponding to different optimal strategies $\mathbf{v} \in \mathcal{B}^n$, and the solution may fail to be twice continuously differentiable across these interfaces.

Remark 1.4. *It is important to point out that there are certainly some limitations to our numerical results. In particular, we cannot solve the PDE (1.1) numerically on the full space $\mathbb{R}^n$ and must restrict*

https://doi.org/10.1017/S0956792525000075 Published online by Cambridge University Press

our attention to a compact subset. Our numerical results are obtained over the box $[/! - 1, 1]^n$. *We cannot rule out, for example, that Conjecture 1.2 fails somewhere outside of the box* $[/! - 1, 1]^n$. *However, we do find that for* $n = 2, 3, 4$ *experts, the optimality observed on the box matches perfectly with the global theory. The negative results of Conjectures 1.1 and 1.3 do not suffer the same limitations, since non-optimality need only be observed at a single point. Finally, our numerical convergence rates rely on classical regularity of the viscosity solution, which is only known for* $n \leq 4$ *experts.*

Remark 1.5. *We also remark that it may be possible to use our techniques for reducing the computational grid to the sector* $\mathbb{S}_n$ *in combination with the dynamic programming approaches of [17, 27]. We leave this interesting direction to future work.*

1.3 Outline

The rest of the paper is organized as follows. In Sections 2 and 3, we present and analyse our numerical scheme for solving the prediction PDE (1.1). The first part in Section 2 is written for a more general class of degenerate elliptic PDEs, while the second part Section 3 contains results that require the specific form of our prediction with expert advice equation. Finally in Section 4, we present the results of our numerical experiments that provide the evidence for the conjectures given in Section 1.2.

2 Analysis of a general numerical scheme

We study here a finite difference scheme for the PDE (1.3) consisting of replacing the pure second derivatives $\mathbf{v}^T\nabla^2 u\mathbf{v}$ with finite difference approximations. We work on the grid $\mathbb{Z}_h^n$, where $\mathbb{Z}_h = h\mathbb{Z}$ and $h > 0$ is the grid spacing. For a function $u{:}\mathbb{Z}_h^n \to \mathbb{R}$, we define the discrete gradient as the mapping $\nabla_h{:}\mathbb{Z}_h^n \times \mathbb{Z}^n \to \mathbb{R}$ defined by

$$\nabla_h u(x, \mathbf{v}) := \frac{u(x + h\mathbf{v}) - u(x)}{h}. \tag{2.1}$$

The discrete Hessian is defined as the mapping $\nabla_h^2 u{:}\mathbb{Z}_h^n \times \mathbb{Z}^n \to \mathbb{R}$ given by

$$\nabla_h^2 u(x, \mathbf{v}) := \frac{u(x + h\mathbf{v}) - 2u(x) + u(x - h\mathbf{v})}{h^2}. \tag{2.2}$$

We note that

$$\nabla_h^2 u(x, \mathbf{v}) = \frac{\nabla_h u(x, \mathbf{v}) + \nabla_h u(x, -\mathbf{v})}{h}. \tag{2.3}$$

Also, the definition of the discrete Hessian leads immediately to the discrete Taylor-type expansions

$$\frac{1}{2}(u(x + h\mathbf{v}) + u(x - h\mathbf{v})) = u(x) + \frac{h^2}{2}\nabla_h^2 u(x, \mathbf{v}), \tag{2.4}$$

and

$$u(x + h\mathbf{v}) = u(x) - h\nabla_h(u, -\mathbf{v}) + h^2\nabla_h^2 u(x, \mathbf{v}). \tag{2.5}$$

Define $\mathcal{H}_n = \{X{:}\mathbb{Z}_n \to \mathbb{R}\}$. We will use the notation $\nabla_h u(x) \in \mathcal{H}_n$ and $\nabla_h^2 u(x) \in \mathcal{H}_n$ for the mappings $\mathbf{v} \mapsto \nabla_h u(x, \mathbf{v})$ and $\mathbf{v} \mapsto \nabla^2 u_h(x, \mathbf{v})$, respectively. For $u \in C^{k,1}(\mathbb{R}^n)$ with $k = 2$ or $k = 3$ and any $\mathbf{v} \in \mathbb{Z}^n$ we have via Taylor expansion that

$$\nabla_h^2 u(x, \mathbf{v}) = \mathbf{v}^T\nabla^2 u(x)\mathbf{v} + \mathcal{O}(|\mathbf{v}|^{k+1}h^{k-1}).$$

Our discrete approximation of (1.3) is given by

$$u(x) - \frac{1}{2}\max_{\mathbf{v}\in\mathcal{B}^n} \nabla_h^2 u(x, \mathbf{v}) = g(x) \quad \text{for } x \in \mathbb{Z}_h^n.$$

https://doi.org/10.1017/S0956792525000075 Published online by Cambridge University Press

Since our methods are not specific to this equation, we will study a more general class of equations of the form

$$u - F(\nabla_h^2 u) = g \ \text{ on } \mathbb{Z}_h^n, \tag{2.6}$$

where $F : \mathcal{H}_n \to \mathbb{R}$. The PDE (1.3) is obtained by setting

$$F(X) = \frac{1}{2} \max_{\mathbf{v} \in \mathcal{B}^n} X(\mathbf{v}). \tag{2.7}$$

We need to place a monotonicity assumption on F.

Definition 2.1. *We say that F:$\mathcal{H}_n \to \mathbb{R}$ is monotone if for all $X, Y \in \mathcal{H}_n$ with $X \leq Y$ we have $F(X) \leq F(Y)$.*

We note that X, Y:$\mathcal{H}_n \to \mathbb{R}$ are real-valued functions, so $X \leq Y$ means that $X(\mathbf{v}) \leq Y(\mathbf{v})$ for all $\mathbf{v} \in \mathbb{Z}_n$. The class of monotone equations of the form (2.6) is closely related to the wide stencil finite difference schemes introduced and studied by Oberman [38]. The main difference in this section is that we are focused on the unbounded domain $\mathbb{Z}_h^n$, and we are interested in properties of the solution u, such as Lipschitzness, convexity, permutation invariance, etc., that hold under certain structure conditions and the source term g.

For $X \in \mathcal{H}_n$ we define

$$\|X\|_{N,\infty} = \max_{\substack{\mathbf{v} \in \mathbb{Z}^n \\ |\mathbf{v}|_\infty \leq N}} |X(\mathbf{v})|, \tag{2.8}$$

where $|\mathbf{v}|_\infty = \max_{1 \leq i \leq n} |v_i|$. We need to place a condition on the width of the stencil F.

Definition 2.2. *We say F:$\mathcal{H}_n \to \mathbb{R}$ has width N if there exists $C > 0$ such that for all $X, Y \in \mathcal{H}_n$ we have*

$$|F(X) - F(Y)| \leq C\|X - Y\|_{N,\infty}.$$

The smallest such constant $C > 0$ is called the Lipschitz constant of F and denoted $Lip_N(F)$.

The choice of F given in (2.7) is monotone and has width $N = 1$, with $\mathrm{Lip}_1(F) = \frac{1}{2}$.

2.1 Existence and uniqueness

We first establish a comparison principle.

Theorem 2.3. *Assume F is monotone and has width N. Let u, v:$\mathbb{Z}_h^n \to \mathbb{R}$ satisfy*

$$u - F(\nabla_h^2 u) \leq v - F(\nabla_h^2 v) \ \text{ on } \mathbb{Z}_h^n, \tag{2.9}$$

and

$$\lim_{|x| \to \infty} \frac{u(x) - v(x)}{|x|^2} = 0. \tag{2.10}$$

Then $u \leq v$ on $\mathbb{Z}_h^n$.

Proof. *For $\varepsilon > 0$ we define*

$$u_\varepsilon(x) = u(x) - \frac{\varepsilon}{2}|x|^2 - \varepsilon N^2 - \varepsilon.$$

We claim that $u_\varepsilon \leq v$ for all $\varepsilon > 0$, from which the result follows. Fix $\varepsilon > 0$ and assume by way of contradiction that $\sup_{\mathbb{Z}_h^n}(u_\varepsilon - v) > 0$. By (2.10), there exists $R > 0$, depending on $\varepsilon > 0$, such that $u_\varepsilon(x) < v(x)$ for $|x| > R$. Thus, $u_\varepsilon - v$ attains its maximum over $\mathbb{Z}_h^n$ at some $x_0 \in \mathbb{Z}_h^n$, and $u_\varepsilon(x_0) > v(x_0)$. Since $u_\varepsilon(x) - v(x) \leq u_\varepsilon(x_0) - v(x_0)$ for all $x \in \mathbb{Z}_h^n$ we have

$$u_\varepsilon(x) - u_\varepsilon(x_0) \leq v(x) - v(x_0).$$

https://doi.org/10.1017/S0956792525000075 Published online by Cambridge University Press

It follows that

$$\nabla_h^2 u_\varepsilon(x_0, \mathbf{v}) = \frac{1}{h^2}(u_\varepsilon(x_0 + h\mathbf{v}) - u(x_0) + u_\varepsilon(x_0 - h\mathbf{v}) - u(x_0))$$
$$\leq \frac{1}{h^2}(v(x_0 + h\mathbf{v}) - v(x_0) + v(x_0 - h\mathbf{v}) - v(x_0)) = \nabla_h^2 v(x_0, \mathbf{v})$$

for all directions $\mathbf{v} \in \mathbb{Z}^n$. *Since* F *is monotone we have*

$$F(\nabla_h^2 u_\varepsilon(x_0)) \leq F(\nabla_h^2 v(x_0)). \tag{2.11}$$

We now compute

$$\nabla_h^2 u_\varepsilon(x, \mathbf{v}) = \nabla_h^2 u(x, \mathbf{v}) - \varepsilon|\mathbf{v}|^2,$$

from which it follows that

$$u_\varepsilon(x_0) - F(\nabla_h^2 u_\varepsilon(x_0)) = u(x_0) - \frac{\varepsilon}{2}|x_0|^2 - \varepsilon N^2 - \varepsilon - F(\nabla_h^2 u(x_0) - \varepsilon|\mathbf{v}|^2)$$
$$\leq u(x_0) - F(\nabla_h^2 u(x_0)) - \frac{\varepsilon}{2}|x_0|^2 - \varepsilon$$
$$< v(x_0) - F(\nabla_h^2 v(x_0))$$
$$\leq v(x_0) - F(\nabla_h^2 u_\varepsilon(x_0)),$$

where the last line follows from (2.11). Therefore $u_\varepsilon(x_0) \leq v(x_0)$, *which is a contradiction.* □

Existence of a solution follows from the comparison principle and the Perron method.

Theorem 2.4. *Assume* F *is monotone, has width* N, *and* $F(0) = 0$. *Suppose there exists* $C_g > 0$ *so that*

$$|g(x)| \leq C_g(1 + |x|) \ \textit{for all } x \in \mathbb{Z}_h^n. \tag{2.12}$$

Then there exists a unique solution $u{:}\mathbb{Z}_h^n \to \mathbb{R}$ *of (2.6) satisfying* $\lim_{|x|\to\infty} \frac{u(x)}{|x|^2} = 0$. *Furthermore, we have*

$$|u(x)| \leq CC_g(Lip_N(F)N^3 + 1 + |x|) \ \textit{for all } x \in \mathbb{Z}_h^n, \tag{2.13}$$

where C *depends only on* n.

Proof. *We define*

$$\psi(x) = \sqrt{1 + |x|^2},$$

and note that ψ *is a smooth function with linear growth satisfying*

$$\frac{1}{\sqrt{2}}(1 + |x|) \leq \psi(x) \leq 1 + |x| \ \textit{for all } x \in \mathbb{R}^n. \tag{2.14}$$

For any $\mathbf{v} \in \mathbb{Z}^n$ *and* $x \in \mathbb{Z}_h^n$, *we have*

$$|\nabla_h^2 \psi(x, \mathbf{v})| \leq |\nabla_h^2 \psi(x, \mathbf{v}) - \mathbf{v}^T \nabla^2 \psi(x)\mathbf{v}| + C|\mathbf{v}|^2$$
$$\leq |\nabla_h^2 \psi(x, \mathbf{v}) - \mathbf{v}^T \nabla^2 \psi(x)\mathbf{v}| + C|\mathbf{v}|^2$$
$$\leq C\left(|\mathbf{v}|^3 h + |\mathbf{v}|^2\right) \leq C|\mathbf{v}|^3,$$

where we used a Taylor series expansion of $\nabla_h^2 \psi(x, \mathbf{v})$ *in the last line, along with* $h \leq 1$ *and* $|\mathbf{v}| \geq 1$. *Since* $F(0) = 0$ *we have*

$$|F(\nabla_h^2 \psi)| = |F(\nabla_h^2 \psi) - F(0)| \leq Lip_N(F)\|\nabla_h^2 \psi\|_{N,\infty} \leq CLip_N(F)N^3 =: \xi. \tag{2.15}$$

We now define

$$w = C_g(\xi + \sqrt{2}\psi).$$

Then by (2.14) and (2.15), we have

$$w - F(\nabla_h^2 w) = C_g\xi + \sqrt{2}C_g\psi - F(C_g\nabla_h^2 \psi) \geq \sqrt{2}C_g\psi \geq C_g(1 + |x|) \geq g.$$

https://doi.org/10.1017/S0956792525000075 Published online by Cambridge University Press

A similar argument shows that $v = -w$ satisfies $v - F(\nabla_h^2 v) \leq g$.

Define

$$\mathcal{F} = \left\{ v : \mathbb{Z}_h^n \to \mathbb{R} : v - F(\nabla_h^2 v) \leq g \text{ and } v \leq w \right\} \tag{2.16}$$

and

$$u(x) = \sup\{v(x) : v \in \mathcal{F}\}. \tag{2.17}$$

Since $v = -w$ belongs to $\mathcal{F}$, the set $\mathcal{F}$ is nonempty and we have $-w \leq u \leq w$. Hence, u satisfies (2.13).

We now claim that

$$u - F(\nabla_h^2 u) \leq g \quad \text{on } \mathbb{Z}_h.$$

To see this, fix $x_0 \in \mathbb{Z}_h^n$ and let $v_k \in \mathcal{F}$ such that $\lim_{k\to\infty} v_k(x_0) = u(x_0)$. By passing to a subsequence, if necessary, we can assume that $\lim_{k\to\infty} v_k(x)$ exists for all $x \in \mathbb{Z}_h^n$. Let us denote $v(x) := \lim_{k\to\infty} v_k(x)$, noting that $v(x_0) = u(x_0)$. By continuity we have $v - F(\nabla_h^2 v) \leq g$ and $v \leq w$, thus $v \in \mathcal{F}$ and $v \leq u$. Since $v(x_0) = u(x_0)$, we have that $v - u$ attains its maximum at x_0. As in the proof of Theorem 2.3 we have $\nabla_h^2 v(x_0, \mathbf{v}) \leq \nabla_h^2 u(x_0, \mathbf{v})$ for all $\mathbf{v}$, and so by the monotonicity of F we have $F(\nabla_h^2 v(x_0)) \leq F(\nabla_h^2 u(x_0))$, which when combined with $u(x_0) = v(x_0)$ and $v - F(\nabla_h^2 v) \leq g$, establishes the claim.

To complete the proof, we show that

$$u - F(\nabla_h^2 u) \geq g \quad \text{on } \mathbb{Z}_h^n.$$

Assume to the contrary that there is some $x_0 \in \mathbb{Z}_h^n$ such that

$$u(x_0) - F(\nabla_h^2 u(x_0)) < g(x_0).$$

Define

$$v(x) = \begin{cases} u(x_0) + \varepsilon, & \text{if } x = x_0 \\ u(x), & \text{otherwise.} \end{cases}$$

By continuity, we can choose $\varepsilon > 0$ small enough so that

$$v(x_0) - F(\nabla_h^2 v(x_0)) \leq g(x_0).$$

For $x \neq x_0$, we already have

$$v(x) - F(\nabla_h^2 v(x)) \leq g(x),$$

and the definition of v and the monotonicity of F imply that $\nabla^2 v(x, \mathbf{v}) \geq \nabla_h^2 u(x, \mathbf{v})$ for all $x \neq x_0$ and $\mathbf{v} \in \mathbb{Z}^n$. This completes the proof. □

2.2 Properties of solutions

We recall that a function $u : \mathbb{Z}_h^n \to \mathbb{R}$ is *Lipschitz continuous* if there exists $C > 0$ such that

$$|u(x) - u(y)| \leq C\|x - y\|$$

holds for all $x, y \in \mathbb{Z}_h^n$. The *Lipschitz constant* of u, denoted $\mathrm{Lip}(u)$, is the smallest such constant, given by

$$\mathrm{Lip}(u) = \sup_{\substack{x,y \in \mathbb{Z}_h^n \\ x \neq y}} \frac{|u(x) - u(y)|}{\|x - y\|}.$$

Lemma 2.5 (Basic properties). *Assume F is monotone, has width N, and satisfies $F(0) = 0$. Assume g satisfies (2.12) and let $u : \mathbb{Z}_h^n \to \mathbb{R}$ be the unique solution of (2.6). The following hold.*

(i) If g is Lipschitz then so is u, and $Lip(u) \leq Lip(g)$.

(ii) If $F \geq 0$ then $u \geq g$ on $\mathbb{Z}_h^n$.

https://doi.org/10.1017/S0956792525000075 Published online by Cambridge University Press

(iii) There exists a constant $C > 0$ such that

$$\|u - g\|_\infty \le C Lip(g) \left(N\sqrt{Lip_N(F)} + h \right). \tag{2.18}$$

Proof. *(i) Let $z \in \mathbb{Z}_h^n$ and define $w(x) = u(x+z) - Lip(g)\|z\|$. Then we have*

$$w(x) - F(\nabla_h^2 w(x)) = u(x+z) - Lip(g)\|z\| - F(\nabla_h^2 u(x+z)) = g(x+z) - Lip(g) \le g(x).$$

By the comparison principle, Theorem 2.3, we have $w \le u$, and so

$$u(x+z) - u(x) \le Lip(g)\|z\|.$$

Since this holds for all $x, z \in \mathbb{Z}_h^n$, the proof of (i) is complete.

(ii) If $F \ge 0$ then $u(x) = g(x) + F(\nabla_h^2 u(x)) \ge g(x)$.

(iii) Let $\overline{g}:\mathbb{R}^n \to \mathbb{R}$ be the piecewise constant extension of g to a function on $\mathbb{R}^n$, defined so that $\overline{g}(y) = g(x)$ for all $y \in x + [0,h)^n$ and any $x \in Zh$. While the extension is discontinuous, it satisfies

$$|\nabla_h \overline{g}(y, \mathbf{v})| = \left| \frac{g(x + h\mathbf{v}) - g(x)}{h} \right| \le Lip(g)|\mathbf{v}|$$

provided $\mathbf{v} \in \mathbb{Z}^n$, where $x \in \mathbb{Z}_h^n$ satisfies $y \in x + [0,h)^n$. Let $\varepsilon > 0$ and define the standard mollification $g_\varepsilon := \eta_\varepsilon * \overline{g}$, where $\eta_\varepsilon(x) = \frac{1}{\varepsilon^d}\eta\left(\frac{x}{\varepsilon}\right)$, and η_ε is compactly supported in $B(0,\varepsilon)$. Note that

$$\nabla_h g_\varepsilon(x, \mathbf{v}) = \int_{\mathbb{R}^n} \eta_\varepsilon(y) \nabla_h \overline{g}(x - y, \mathbf{v})\, dy = \int_{\mathbb{R}^n} \eta_\varepsilon(x - z) \nabla_h \overline{g}(z, \mathbf{v})\, dz,$$

where we used the change of variables $z = x - y$ above. We also have

$$\nabla_h g_\varepsilon(x, -\mathbf{v}) = \int_{\mathbb{R}^n} \eta_\varepsilon(y) \nabla_h \overline{g}(x - y, -\mathbf{v})\, dy = -\int_{\mathbb{R}^n} \eta_\varepsilon(x - z - h\mathbf{v}) \nabla_h \overline{g}(z, \mathbf{v})\, dz,$$

where we now use the change of variables $z = x - y - h\mathbf{v}$. Therefore

$$\begin{aligned} \nabla_h^2 g_\varepsilon(x, \mathbf{v}) &= \frac{1}{h}(\nabla_h \overline{g}_\varepsilon(x, \mathbf{v}) + \nabla_h \overline{g}_\varepsilon(x, -v)) \\ &= -\int_{\mathbb{R}^n} \nabla_h \eta_\varepsilon(x - z, -\mathbf{v}) \nabla_h \overline{g}(z, \mathbf{v})\, dz. \end{aligned}$$

Therefore,

$$|\nabla_h^2 g_\varepsilon(x, \mathbf{v})| \le \int_{B(x,\varepsilon) \cup B(x - h\mathbf{v}, \varepsilon)} |\nabla_h \eta_\varepsilon(x - z, -\mathbf{v})| |\nabla_h \overline{g}(z, \mathbf{v})|\, dz \le \frac{C}{\varepsilon} Lip(g)|\mathbf{v}|^2.$$

Since F has width N, it follows that

$$|F(\nabla_h^2 g_\varepsilon(x))| = |F(\nabla_h^2 g_\varepsilon(x)) - F(0)| \le Lip_N(F)\|\nabla_h^2 g_\varepsilon\|_{N,\varepsilon} \le \frac{C}{\varepsilon} Lip_N(F) Lip(g) N^2. \tag{2.19}$$

We also have that $|\overline{g}(y) - g(x)| \le |y - x| + \sqrt{d}h$, and so

$$\begin{aligned} |g_\varepsilon(x) - g(x)| &= \left| \int_{B(x,\varepsilon)} \eta_\varepsilon(x - y)(\overline{g}(y) - g(x))\, dy \right| \\ &\le \int_{B(x,\varepsilon)} \eta_\varepsilon(x - y)|\overline{g}(y) - g(x)|\, dy \\ &\le Lip(g)(\varepsilon + \sqrt{d}h). \end{aligned}$$

Combining this with (2.19) we have

$$|g_\varepsilon(x) + F(\nabla^2 g_\varepsilon(x)) - g(x)| \le Lip(g)(\varepsilon + \sqrt{d}h) + \frac{C}{\varepsilon} Lip_N(F) Lip(g) N^2.$$

Choosing $\varepsilon = N\sqrt{Lip_N(F)}$ yields

$$|g_\varepsilon(x) + F(\nabla^2 g_\varepsilon(x)) - g(x)| \le C Lip(g) \left(N\sqrt{Lip_N(F)} + h \right).$$

https://doi.org/10.1017/S0956792525000075 Published online by Cambridge University Press

By the comparison principle, Theorem 2.3, we have

$$g_\varepsilon - CLip(g)\left(N\sqrt{Lip_N(F)}+h\right) \le u \le g_\varepsilon + CLip(g)\left(N\sqrt{Lip_N(F)}+h\right),$$

which completes the proof. □

To study further properties of u, we need some additional definitions.

Definition 2.6. *We say that $F:\mathcal{H}_n \to \mathbb{R}$ is convex if for all $X, Y \in \mathcal{H}_n$ and $\lambda \in [0,1]$ we have*

$$F(\lambda X + (1-\lambda)Y) \le \lambda F(X) + (1-\lambda)F(Y).$$

Definition 2.7. *We say that $u:\mathbb{Z}_h^n \to \mathbb{R}$ is convex if $\nabla_h^2 u(x) \ge 0$ for all $x \in \mathbb{Z}_h^n$.*

We note that by (2.4), the convexity of u is equivalent to the inequality

$$u(x) \le \frac{1}{2}(u(x+h\mathbf{v}) + u(x-h\mathbf{v}))$$

holding for all $x \in \mathbb{Z}_h^n$ and $\mathbf{v} \in \mathbb{Z}^n$. We also note that the choice of F given in (2.7) is convex.

We also consider a permutation invariance property.

Definition 2.8. *We say $u:\mathbb{Z}_h^n \to \mathbb{R}$ is permutation invariant if $u \circ \sigma = u$ for all permutations σ on $\{1,\dots,n\}$.*

Definition 2.9. *We say $F:\mathcal{H}_n \to \mathbb{R}$ is permutation invariant if $F(X \circ \sigma) = F(X)$ for all $X \in \mathcal{H}_n$ and all permutations σ on $\{1,\dots,n\}$.*

We note that F given in (2.7) is permutation invariant.

Finally, we also study translation properties.

Definition 2.10. *For $\mathbf{v} \in \mathbb{Z}^n$, we say $u:\mathbb{Z}_h^n \to \mathbb{R}$ satisfies the $\mathbf{v}$-translation property if there exists a constant $c_\mathbf{v}$ such that*

$$u(x+s\mathbf{v}) = u(x) + c_\mathbf{v} s \tag{2.20}$$

for all $x \in \mathbb{Z}_h^n$ and $s \in \mathbb{R}$ such that $s\mathbf{v} \in \mathbb{Z}_h^n$.

The next lemma shows that these properties for u are inherited from F and g.

Lemma 2.11. *Assume F is monotone, has width N, and satisfies $F(0) = 0$. Assume g satisfies (2.12) and let $u:\mathbb{Z}_h^n \to \mathbb{R}$ denote the unique solution of (2.6). The following hold:*

(i) If F and g are convex, then u is convex.

(ii) If F and g are permutation invariant, then u is permutation invariant.

(iii) If g satisfies the $\mathbf{v}$-translation property with constant $c_\mathbf{v}$, then u does as well.

Proof. *(i) Let $\mathbf{v} \in \mathbb{Z}^n$ and define*

$$w(x) = \frac{1}{2}(u(x+h\mathbf{v}) + u(x-h\mathbf{v})).$$

https://doi.org/10.1017/S0956792525000075 Published online by Cambridge University Press

Then for $x \in \mathbb{Z}_h^n$ we compute, using the convexity of F, that

$$\begin{aligned} w(x) - F(\nabla_h^2 w(x)) &= \frac{1}{2}u(x+h\mathbf{v}) + \frac{1}{2}u(x-h\mathbf{v}) - F\left(\frac{1}{2}\nabla_h^2 u(x+h\mathbf{v}) + \frac{1}{2}\nabla_h^2 u(x-h\mathbf{v})\right) \\ &\geq \frac{1}{2}u(x+h\mathbf{v}) + \frac{1}{2}u(x-h\mathbf{v}) - \frac{1}{2}F(\nabla_h^2 u(x+h\mathbf{v})) - \frac{1}{2}F(\nabla_h^2 u(x-h\mathbf{v})) \\ &= \frac{1}{2}g(x+h\mathbf{v}) + \frac{1}{2}g(x-h\mathbf{v}) \\ &= \frac{h^2}{2}\nabla_h^2 g(x,\mathbf{v}) + g(x) \\ &\geq g(x), \end{aligned}$$

where the last line follows from the convexity of g. By Theorem 2.3 we have $w \geq u$, and so

$$\nabla_h^2 u(x,\mathbf{v}) = \frac{2}{h^2}(w(x) - u(x)) \geq 0$$

for all $x \in \mathbb{Z}_h^n$. Since $\mathbf{v} \in \mathbb{Z}^n$ is arbitrary, we have $\nabla_h^2 u(x) \geq 0$ for all $x \in \mathbb{Z}^n$, hence u is convex.

(ii) Let σ be a permutation and set $w = u \circ \sigma$. Then we have

$$w(x) - F(\nabla_h^2 w(x)) = u(\sigma(x)) - F(\nabla_h^2 (u \circ \sigma)(x)).$$

We note that for any $\mathbf{v} \in \mathbb{Z}^n$ we have $\nabla_h^2 (u \circ \sigma)(x,\mathbf{v}) = \nabla_h^2 u(\sigma(x), \sigma(\mathbf{v}))$. Therefore $\nabla_h^2 (u \circ \sigma)(x) = \nabla_h^2 u(\sigma(x)) \circ \sigma$. Since F is permutation invariant we have

$$F(\nabla_h^2 (u \circ \sigma)(x)) = F(\nabla_h^2 u(\sigma(x)) \circ \sigma) = F(\nabla_h^2 u(\sigma(x))).$$

Therefore,

$$w(x) - F(\nabla_h^2 w(x)) = u(\sigma(x)) - F(\nabla_h^2 u(\sigma(x))) = 0.$$

By uniqueness of the solution u of (2.6), we have $w = u$, which completes the proof.

(iii) Let $s \in \mathbb{R}$ such that $s\mathbf{v} \in \mathbb{Z}_h^n$, and define $w(x) = u(x + s\mathbf{v}) - c_{\mathbf{v}} s$. Then,

$$w(x) - F(\nabla_h^2 w(x)) = u(x+s\mathbf{v}) - c_{\mathbf{v}} s - F(\nabla_h^2 u(x+s\mathbf{v})) = g(x+s\mathbf{v}) - c_{\mathbf{v}} s = g(x).$$

Since the solution of (2.6) is unique, we have $w = u$, which completes the proof. □

2.3 Convergence rates

In this section, we prove convergence of the numerical scheme (2.6) towards the viscosity solution of the second-order degenerate elliptic equation

$$u - F(\nabla^2 u) = g \quad \text{on } \mathbb{R}^n. \tag{2.21}$$

As before, we assume $F: \mathcal{H}_n \to \mathbb{R}$, and we interpret $F(X)$ for an $n \times n$ symmetric matrix X as

$$F(X) := F(\mathbf{v} \mapsto \mathbf{v}^T X \mathbf{v}).$$

We recall the definitions of viscosity solutions in Appendix A. Throughout this section, we will use the notation USC($\mathcal{O}$) (resp. LSC($\mathcal{O}$)) for the set of functions that are upper (resp. lower) semicontinuous at all points in $\mathcal{O} \subset \mathbb{R}^n$. For more details on the theory of viscosity solutions, we refer the reader to the user's guide [19] and [10].

Existence and uniqueness of a linear growth viscosity solution to (2.21) is standard material for viscosity solutions, and proofs can be found in [10, 19]. For use later on, we recall the comparison principle for (2.21) in Lemma 2.12 below. The proof of this result is standard in viscosity solution theory; a self-contained proof can be found in [10, Lemma 12.17].

Lemma 2.12. *Assume F is uniformly continuous, monotone, has width N, and satisfies $F(0) = 0$. Let $u \in USC(\mathbb{R}^n)$ be a viscosity subsolution of (2.21) and let $v \in LSC(\mathbb{R}^n)$ be a viscosity*

https://doi.org/10.1017/S0956792525000075 Published online by Cambridge University Press

supersolution of (2.21). If

$$\lim_{|x|\to\infty} \frac{u(x)-v(x)}{|x|^2} = 0 \tag{2.22}$$

then $u \leq v$ on $\mathbb{R}^n$.

Using this comparison principle and the Perron method, we can prove existence of a linear growth solution (Theorem 2.13 below). The application of the Perron method is standard and a self-contained proof can be found in [10, Theorem 12.18].[2]

Theorem 2.13. *Assume F is monotone, has width N, and satisfies $F(0) = 0$. Assume g is Lipschitz continuous and there exists $C_g > 0$ such that*

$$|g(x)| \leq C_g(1+|x|). \tag{2.23}$$

Then there exists a unique viscosity solution $u \in C(\mathbb{R}^n)$ of (2.21) satisfying

$$\lim_{|x|\to\infty} \frac{u(x)}{|x|^2} = 0.$$

Furthermore, there exists $C > 0$ such that

$$|u(x)| \leq C(1+|x|). \tag{2.24}$$

Convergence of the discrete scheme (2.6) to the PDE (2.21) is a standard result in viscosity solution theory. We state the theorem in the following result and briefly sketch the proof.

Theorem 2.14. *Assume F is monotone, has width N, and satisfies $F(0) = 0$. Assume $g{:}\mathbb{R}^n \to \mathbb{R}$ is Lipschitz continuous and satisfies (2.23). Let u_h be the solution of (2.6) and let u be the viscosity solution of (2.21). Then u_h converges to u locally uniformly as $h \to 0$, that is for all $R > 0$ we have*

$$\lim_{h\to 0} \max_{\substack{x\in\mathbb{Z}_h^n \\ |x|\leq R}} |u_h(x) - u(x)| = 0. \tag{2.25}$$

Proof. *By Lemma 2.5 (i), we have $Lip(u_h) \leq Lip(g)$. By the Arzelá-Ascoli Theorem there exists a Lipschitz continuous function $u{:}\mathbb{R}^n \to \mathbb{R}$ such that, upon passing to a subsequence u_{h_k}, we have*

$$\lim_{k\to\infty} \max_{\substack{x\in\mathbb{Z}_{h_k}^n \\ |x|\leq R}} |u_{h_k}(x) - u(x)| = 0$$

for all $R > 0$. The proof will be completed by showing that u is a viscosity solution of (2.21). By uniqueness of viscosity solutions of (2.21), the whole sequence u_h converges locally uniformly to u.

We verify the subsolution property for u; the supersolution property is similar. Let $x_0 \in \mathbb{R}^n$ and $\varphi \in C^\infty(\mathbb{R}^n)$ such that $u - \varphi$ has a local maximum at x_0. Without loss of generality, we can assume x_0 is a strict global maximum of $u - \varphi$. It follows that there exists $x_k \to x_0$ such that $u_{h_k} - \varphi$ attains its maximum over $\mathbb{Z}_{h_k}^n$ at x_k. Therefore $\nabla_h^2 u_{h_k}(x_k, \mathbf{v}) \leq \nabla_h^2 \varphi(x_k)$, and since F is monotone we have

$$u(x_k) - F(\nabla_{h_k}^2 \varphi(x_k)) \leq \varphi(x_k) - F(\nabla_h^2 u_{h_k}(x_k)) = u(x_k) - u_{h_k}(x_k) - g(x_k).$$

Sending $k \to \infty$ we obtain

$$u(x_0) - F(\nabla^2 \varphi(x_0)) \leq g(x_0),$$

which completes the proof. □

If we have additional regularity for the solution u of (2.21), we can prove convergence rates. We recall the $C^{k,1}$ seminorm of $u{:}\mathbb{R}^n \to \mathbb{R}$ is defined by

$$[u]_{C^{k,1}(\mathbb{R}^n)} = \sum_{1\leq|\alpha|\leq k} \mathrm{Lip}(D^\alpha u).$$

[2] In fact, the proofs do not require Lipschitzness; uniform continuity is sufficient.

https://doi.org/10.1017/S0956792525000075 Published online by Cambridge University Press

Theorem 2.15. *Assume F is monotone, has width N, and satisfies $F(0) = 0$. Assume $g:\mathbb{R}^n \to \mathbb{R}$ is Lipschitz continuous and satisfies (2.23). Let u_h be the solution of (2.6) and let u be the viscosity solution of (2.6). If $[u]_{C^{k,1}(\mathbb{R}^n)} < \infty$ for $k = 2$ or $k = 3$, then we have*

$$\|u - u_h\|_\infty \leq CLip_N(F)[u]_{C^{k,1}(\mathbb{R}^n)} N^{k+1} h^{k-1} \tag{2.26}$$

Proof. *By Taylor expansion, we have*

$$|\nabla_h^2 u(x, \mathbf{v}) - \mathbf{v}^T \nabla^2 u(x) \mathbf{v}| \leq C[u]_{C^{k,1}(\mathbb{R}^n)} |\mathbf{v}|^{k+1} h^{k-1}.$$

Therefore,

$$|F(\nabla_h^2 u(x)) - F(\nabla^2 u(x))| \leq CLip_N(F)[u]_{C^{k,1}(\mathbb{R}^n)} N^{k+1} h^{k-1} =: \varepsilon.$$

It follows that

$$u(x) - F(\nabla_h^2 u(x)) \leq u(x) - F(\nabla^2 u(x)) + \varepsilon = g(x) + \varepsilon \quad \text{for } x \in \mathbb{Z}_h^n.$$

By Theorem 2.3, we have $u \leq u_h + \varepsilon$ on $\mathbb{Z}_h^n$. The opposite inequality is obtained similarly. □

From the convergence results in Theorems 2.14 and 2.15, we immediately obtain that all the properties of the discrete solutions proved in Section 2.2 extend to the viscosity solution u.

Proposition 2.16. *Assume F is monotone, has width N, and satisfies $F(0) = 0$. Assume $g:\mathbb{R}^n \to \mathbb{R}$ is Lipschitz continuous and satisfies (2.23). Let u be the viscosity solution of (2.6). Then the following hold.*

(i) *u is Lipschitz continuous and $Lip(u) \leq Lip(g)$.*

(ii) *If $F \geq 0$ then $u \geq g$ on $\mathbb{R}^n$.*

(iii) *There exists a constant $C > 0$ such that*

$$\|u - g\|_\infty \leq CNLip(g)\sqrt{Lip_N(F)}. \tag{2.27}$$

(iv) *If F and g are convex, then u is convex.*

(v) *If F and g are permutation invariant, then u is permutation invariant.*

(vi) *If g satisfies the $\mathbf{v}$-translation property with constant $c_\mathbf{v}$, then u does as well.*

We note that in (iv), the notion of convexity for F is the same as in Definition 2.6, while for u it is the usual one for functions on $\mathbb{R}^n$, that is

$$u(\lambda x + (1 - \lambda)y) \leq \lambda u(x) + (1 - \lambda)u(y).$$

In (v), the definition of permutation invariance for $u:\mathbb{R}^n \to \mathbb{R}$ is identical to Definition 2.8; namely $u = u \circ \sigma$. The translation property is defined similarly, but is slightly different so we give the definition below for functions on $\mathbb{R}^n$.

Definition 2.17. *For $\mathbf{v} \in \mathbb{R}^n$, we say $u:\mathbb{R}^n \to \mathbb{R}$ satisfies the $\mathbf{v}$-translation property if there exists a constant $c_\mathbf{v}$ such that (2.20) holds for all $x \in \mathbb{R}^n$ and $s \in \mathbb{R}$.*

The proof of Proposition 2.16 follows from Lemmas 2.5 and 2.11, and the convergence result in Theorem 2.14.

2.4 Restricting the domain

In order to compute the solution of the discrete scheme (2.6) on an unbounded domain $\mathbb{Z}_h^n$, it is necessary to restrict the domain to a compact set. In this section, we study restrictions of (1.3) to *computational* domains of the form $\Omega_{T,h} := [-T, T]^n \cap \mathbb{Z}_h^n$. In this section, we always assume $T = mh$ for some integer $m \geq 1$. We define the width N boundary as

$$\partial_N \Omega_{T,h} = \{x \in \mathbb{Z}_h^n \setminus \Omega_{T,h} : \text{there exists } y \in \Omega_{T,h} \text{ with } |x - y|_\infty \leq N\}. \tag{2.28}$$

https://doi.org/10.1017/S0956792525000075 Published online by Cambridge University Press

We set a Dirichlet boundary condition on $\partial_N \Omega_{T,h}$ and show that this condition affects the solution only near the boundary, and the solution remains accurate in the interior of $\Omega_{T,h}$.

In this section, we study the equation

$$u - F(\nabla_h^2 u) = g \quad \text{on} \quad \Omega_{T,h}, \tag{2.29}$$

which is the restriction of (2.6) to the computational domains $\Omega_{T,h}$. We first recall the comparison principle for (2.29).

Lemma 2.18. *Assume F is monotone, has width N, and satisfies $F(0) = 0$. Let $u, v : \mathbb{Z}_h^n \to \mathbb{R}$ satisfy*

$$u - F(\nabla_h^2 u) \leq v - F(\nabla_h^2 v) \quad \text{on} \quad \Omega_{T,h} \tag{2.30}$$

and $u \leq v$ on $\partial_N \Omega_{T,h}$. Then $u \leq v$ on $\Omega_{T,h}$.

Proof. *Let $x_0 \in \Omega_{T,h} \cup \partial_N \Omega_{T,h}$ be a point where $u - v$ attains its maximum value over $\Omega_{T,h} \cup \partial_N \Omega_{T,h}$. If $x_0 \in \partial_N \Omega_{T,h}$ then $u \leq v$, so we may assume $x_0 \in \Omega_{T,h}$. In this case we have $\nabla_h^2 u(x_0, \mathbf{v}) \leq \nabla_h^2 v(x_0, \mathbf{v})$, and since F is monotone, we obtain*

$$u(x_0) - v(x_0) \leq F(\nabla_h^2 u(x_0)) - F(\nabla_h^2 v(x_0)) \leq 0.$$

This completes the proof. □

We now establish localization of solutions of (2.29).

Theorem 2.19. *Assume F is monotone, has width N, and satisfies $F(0) = 0$. Let $u, v : \mathbb{Z}_h^n \to \mathbb{R}$ satisfy (2.30). Then for any $\alpha \in (0, 1)$ we have*

$$\max_{\Omega_{\alpha T,h}} (u - v) \leq \frac{C Lip_N(F) N^2}{(1-\alpha)^2 T^2} \left(\log \left(\frac{T^2}{Lip_N(F) N^2} \right)^2 + 1 \right) \max_{\partial_N \Omega_{T,h}} (u - v). \tag{2.31}$$

Remark 2.20. *The estimate (2.31) in Theorem 2.19 shows that the Dirichlet boundary conditions on $\partial_N \Omega_{T,h}$ have a limited domain of influence on the solution in $\Omega_{T,h}$ when $T > 0$ is large. Indeed, if we fix, say, $\alpha = \frac{1}{2}$, and if $u, v : \mathbb{Z}_h^n \to \mathbb{R}$ are solutions of (2.29), then u and v satisfy (2.30) with equality, and so it follows from two applications of Theorem 2.19, applied to $u - v$ and $v - u$, that*

$$\max_{\Omega_{T/2,h}} |u - v| \leq \frac{C \log (T)^2}{T^2} \max_{\partial_N \Omega_{T,h}} |u - v|, \tag{2.32}$$

holds, where C depends on $\text{Lip}_N(F)$ and N, and $T \geq 3$. This shows that errors in the Dirichlet condition on $\partial_N \Omega_{T,h}$ can be tolerated numerically, provided $T > 0$ is large enough. For example, by Lemma 2.5 (iii), we can set $u = g$ on $\partial_N \Omega_{T,h}$ and obtain an $\mathcal{O}(\log (T)^2 / T^2)$ approximation of the solution of (2.21) on the interior domain $\Omega_{T/2,h}$. We show how to obtain even more accurate solutions with better choices of boundary conditions in Section 3.1.

Proof. *Let $\mu = \max_{\partial_N \Omega_{T,h}} (u - v)$ and define*

$$w(x) = v(x) + \gamma + \mu \sum_{i=1}^{n} \left(e^{\lambda \left(\frac{x_i}{T} - 1 \right)} + e^{-\lambda \left(\frac{x_i}{T} + 1 \right)} \right) \tag{2.33}$$

for parameters $\lambda, \gamma > 0$ to be determined. For $x \in \partial_N \Omega_{T,h}$, there exists i such that $x_i \geq T$ or $x_i \leq -T$, and so $w(x) \geq v(x) + \mu \geq u(x)$. Note that we have

$$|F(\nabla_h^2 w) - F(\nabla_h^2 v)| \leq Lip_N(F) \| \nabla_h^2 w - \nabla_h^2 v \|_{N,\infty} \leq C Lip_N(F) \frac{\mu \lambda^2 N^2}{T^2}.$$

Choosing $\gamma = C Lip_N(F) \mu \lambda^2 N^2 T^{-2}$ we find that

$$w(x) - F(\nabla_h^2 w(x)) \geq v(x) + \gamma - F(\nabla_h^2 v(x)) - C Lip_N(F) \mu \lambda^2 N^2 T^{-2} \geq 0.$$

https://doi.org/10.1017/S0956792525000075 Published online by Cambridge University Press

By Lemma 2.18 we have $w \geq u$ *on* $\Omega_{T,h}$, *and so*

$$u(x) - v(x) \leq \gamma + \mu \sum_{i=1}^{n} \left(e^{\lambda\left(\frac{x_i}{T}-1\right)} + e^{-\lambda\left(\frac{x_i}{T}+1\right)} \right)$$

for all $x \in \Omega_{T,h}$. *For* $x \in \Omega_{\alpha T,h}$ *with* $\alpha \in (0,1)$ *we have*

$$u(x) - v(x) \leq \gamma + 2n\mu e^{-\lambda(1-\alpha)}.$$

Choosing λ *so that* $(1-\alpha)\lambda = \log\left(\frac{T^2}{Lip_{N'}(F)N^2}\right)$ *completes the proof.* □

3. Numerical analysis of the prediction PDE

This section is concerned with numerical analysis specific to the prediction from expert advice numerical scheme

$$u(x) - \frac{1}{2} \max_{\mathbf{v} \in \mathcal{B}^n} \nabla_h^2 u(x, \mathbf{v}) = g(x) \quad \text{for } x \in \mathbb{Z}_h^n. \tag{3.1}$$

In particular, we show in Section 3.1 how to use the translation property (2.20) to reduce the dimension by one, and in Section 3.2, we show how to reduce the computational domain to a sector where the coordinates are ordered.

3.1 Reducing the dimension

We show here how to reduce the problem from n dimensional to $n-1$ dimensional. This requires the translation property (2.20) and is based on the following lemma.

Lemma 3.1. *Suppose* u *satisfies the translation property (2.20) with* $\mathbf{v} = \mathbb{1}$ *and* $c_{\mathbf{v}} = 1$. *Then for all* $\mathbf{v} \in \mathcal{B}^n$ *we have*

$$\nabla_h^2 u(x, \mathbf{v}) = \nabla_h^2 u(x, \mathbb{1} - \mathbf{v}). \tag{3.2}$$

Proof. *By assumption we have*

$$u(x + s\mathbb{1}) = u(x) + s \tag{3.3}$$

for all $s \in \mathbb{Z}_h$ *and* $x \in \mathbb{Z}_h^n$. *We now compute, using (3.3), that*

$$\begin{aligned} h^2 \nabla_h^2 u(x, \mathbf{v}) &= u(x + h\mathbf{v}) - 2u(x) + u(x - h\mathbf{v}) \\ &= u(x + h\mathbf{v} - h\mathbb{1}) + h - 2u(x) + u(x - h\mathbf{v} + h\mathbb{1}) - h \\ &= u(x - h(\mathbb{1} - \mathbf{v})) - 2u(x) + u(x + h(\mathbb{1} - \mathbf{v})) \\ &= h^2 \nabla_h^2 u(x, \mathbb{1} - \mathbf{v}), \end{aligned}$$

which completes the proof. □

For $x \in \mathbb{R}^{n-1}$ and $a \in \mathbb{R}$ we define $(x, a) \in \mathbb{R}^n$ by

$$(x, a) = (x_1, x_2, \ldots, x_{n-1}, a).$$

The next lemma shows how we can use the translation property (2.20) to reduce (3.1) to a similar equation in one less variable.

Lemma 3.2. *Let* u *be the solution of (3.1) and suppose* u *satisfies the translation property (2.20) with* $\mathbf{v} = \mathbb{1}$ *and* $c_{\mathbf{v}} = 1$. *Define* $w, f : \mathbb{Z}_h^{n-1} \to \mathbb{R}$ *by* $w(x) := u(x, 0)$ *and* $f(x) = g(x, 0)$. *The following hold.*

(i) The function w *satisfies*

$$w(x) - \frac{1}{2} \max_{\mathbf{v} \in \mathcal{B}^{n-1}} \nabla_h^2 w(x, \mathbf{v}) = f(x) \quad \text{for all} \quad x \in \mathbb{Z}_h^{n-1}. \tag{3.4}$$

https://doi.org/10.1017/S0956792525000075 Published online by Cambridge University Press

(ii) If g is permutation invariant, then so is w, and furthermore w satisfies

$$w(x) = w\left(x - x_i(\mathbb{1} + \mathbf{e}_i)\right) + x_i \quad \text{for all } x \in \mathbb{Z}_h^{n-1} \text{ and } i = 1, \dots, n-1. \tag{3.5}$$

Proof. *By Lemma 3.1, for any $x \in \mathbb{Z}_h^{n-1}$ we have*

$$\max_{\mathbf{v}\in\mathcal{B}^n} \nabla_h^2 u((x,0),\mathbf{v}) = \max_{\mathbf{v}\in\mathcal{B}^{n-1}} \nabla_h^2 u((x,0),(\mathbf{v},0)) = \max_{\mathbf{v}\in\mathcal{B}^{n-1}} \nabla_h^2 w(x,\mathbf{v}).$$

Since u satisfies (3.1) we have

$$w(x) - \frac{1}{2}\max_{\mathbf{v}\in\mathcal{B}^{n-1}} \nabla_h^2 w(x,\mathbf{v}) = g(x,0) = f(x),$$

which completes the proof.

(ii) Since g is permutation invariant, so is u (by Lemma 2.11) and hence w. Let σ_{ij} denote the permutation swapping coordinates i and j. Let $x \in \mathbb{R}^{n-1}$ and let $y = (x,0)$. For any $1 \le i \le n-1$ we use the translation property and permutation invariance to obtain

$$\begin{aligned} w(x) = u(\,(x,0)\,) &= u\left((x,0) - x_i\mathbb{1}\right) + x_i \\ &= u\left((x - x_i\mathbb{1}, -x_i)\right) + x_i \\ &= u\left(\sigma_{i,n}\,(x - x_i\mathbb{1}, -x_i)\right) + x_i \\ &= u\left(x - x_i\mathbb{1} - x_i\mathbf{e}_i, 0\right) + x_i \\ &= w\left(x - x_i\mathbb{1} - x_i\mathbf{e}_i\right) + x_i, \end{aligned}$$

which completes the proof. □

By Lemma 3.2 and Remark 2.20, we may instead solve the equation

$$\begin{cases} w(x) - \dfrac{1}{2}\displaystyle\max_{\mathbf{v}\in\mathcal{B}^{n-1}} \nabla_h^2 w(x,\mathbf{v}) = g(x,0), & \text{if } x \in \Omega_{T,h} \\ w(x) = g(x,0), & \text{if } x \in \partial_1\Omega_{T,h} \end{cases} \tag{3.6}$$

in dimension $n-1$, where we take $T = mh$ to be a multiple of the grid resolution. Provided we restrict our attention to the localized interior set $\Omega_{\alpha T,h}$ for some $\alpha \in (0,1)$, then as per Remark 2.20 we incur an $O(1/T^2)$ error term, up to logarithmic factors.

3.2 Reducing the domain to a sector

It turns out that in addition to reducing the dimension of the equation from n to $n-1$, we can also drastically reduce the size of the computational grid by restricting our attention to the sector

$$\mathbb{D}_n = \{x \in \mathbb{Z}_h^n : x_1 \ge x_2 \ge \cdots \ge x_n\}, \tag{3.7}$$

and the positive sector

$$\mathbb{D}_n^+ = \{x \in \mathbb{D}_n : x_n \ge 0\}. \tag{3.8}$$

Whenever g is permutation invariant, e.g., the max-regret $g(x) = \max\{x_1, \dots, x_n\}$, the solution u of (3.1) and the reduced solution $w(x) = u(x,0)$ are also permutation invariant. As we show in this section, this, combined with the translation property, or (3.5), allows us to reduce the domain of the discrete PDE (3.1) to $\mathbb{D}_n^+$, which is drastically smaller than the full computational grid $\Omega_{T,h}$.

In order to do this in a computational setting, for $x \in \mathbb{D}_{n-1}^+$ and $v \in \mathcal{B}^{n-1}$, we need to be able to evaluate $w(x+\mathbf{v})$ and $w(x-\mathbf{v})$ in terms of only the values of w within the positive sector $\mathbb{D}_{n-1}^+$. This will allow us to evaluate the discrete second derivative $\nabla_h^2 w(x,\mathbf{v})$ without reference to the values of u outside of $\mathbb{D}_+^{n-1}$. To do this, we need to define two operations. First, let $\pi_n : \mathbb{R}^n \to \mathbb{R}^n$ be the *sorting function* that sorts the coordinates of an n-dimensional vectors. Thus, $\pi_n(\mathbb{Z}_h^n) = \mathbb{D}_n$. We also define the function $\xi : \mathbb{D}^n \to \mathbb{R}^n$ by

https://doi.org/10.1017/S0956792525000075 Published online by Cambridge University Press

$$\xi_n(x) = \begin{cases} x, & \text{if } x \in \mathbb{D}_n^+, \\ x + h(\mathbb{1} + \mathbf{e}_i), & \text{where } x_1, \ldots, x_{i-1} \geq 0 > x_i \geq \cdots \geq x_n. \end{cases} \tag{3.9}$$

To indicate the cases above, we write

$$\mathbb{1}_{\xi_n}(x) = \begin{cases} 0, & \text{if } \xi_n(x) = x \\ 1, & \text{otherwise.} \end{cases}$$

The following lemma shows how to evaluate $w(x \pm \mathbf{v})$ within the positive sector $\mathbb{D}_{n-1}^+$.

Lemma 3.3. *Suppose $w{:}\mathbb{Z}_h^{n-1} \to \mathbb{R}$ is permutation invariant and satisfies (3.5). Let $T > 0$, $x \in \mathbb{D}_{n-1}^+ \cap [0, T-h]^{n-1}$ and $\mathbf{v} \in \mathcal{B}^{n-1}$. Then the following hold.*

(i) For $y := \pi_{n-1}(x + h\mathbf{v})$ we have $y \in \mathbb{D}_{n-1}^+ \cap [0, T]^{n-1}$ and $w(x + h\mathbf{v}) = w(y)$.

(ii) For $y := \pi_{n-1}(x - h\mathbf{v})$ we have $\xi_{n-1}(y) \in \mathbb{D}_{n-1}^+ \cap [0, T]^{n-1}$ and

$$w(x - h\mathbf{v}) = w(\xi_{n-1}(y)) - h\mathbb{1}_{\xi_{n-1}}(y).$$

Proof. *Part (i) follows directly from the permutation invariance of w and that $\mathbf{v}$ is a binary vector.*

For part (ii), let $y = \pi_{n-1}(x - h\mathbf{v}) \in \mathbb{D}_{n-1}$. If $y \in \mathbb{D}_{n-1}^+$, then $\xi_{n-1}(y) = y$, $\mathbb{1}_{\xi_{n-1}}(y) = 0$, and the result follows from the permutation invariance of w, as in part (i). If $y \notin \mathbb{D}_{n-1}^+$, then $\xi_{n-1}(y) \neq y$ and $\mathbb{1}_{\xi_{n-1}}(y) = 1$. Since $x \in \mathbb{D}_{n-1}^+$ we must have

$$y_1 \geq \cdots \geq y_{i-1} \geq 0 > -h = y_i = \cdots = y_{n-1},$$

for some $1 \leq i \leq n-1$. Therefore

$$\xi_{n-1}(y) = y + h(\mathbb{1} + \mathbf{e}_i) = y - y_i(\mathbb{1} + \mathbf{e}_i),$$

and so by (3.5) and the permutation invariance of w we have

$$w(x - h\mathbf{v}) = w(y) = w(\xi_{n-1}(y)) + y_i = w(\xi_{n-1}(\pi_{n-1}(x - h\mathbf{v}))) - h,$$

which completes the proof. □

Lemma 3.3 allows us to restrict the reduced $n-1$ dimensional equation for w to the sector $\mathbb{D}_{n-1}^+$, yielding the equation

$$\begin{cases} w(x) - \dfrac{1}{2} \max\limits_{\mathbf{v} \in \mathcal{B}^{n-1}} \nabla_h^2 w(x, \mathbf{v}) = g(x, 0), & \text{if } x \in \mathbb{D}_{n-1}^+ \cap \{x_1 \leq T - h\} \\ w(x) = g(x, 0), & \text{if } x \in \mathbb{D}_{n-1}^+ \cap \{x_1 = T\}, \end{cases} \tag{3.10}$$

where $T = mh$ is a multiple of the grid resolution. By Lemma 3.3 we can compute the derivative $\nabla_h^2 w(x, \mathbf{v})$ for $x \in \mathbb{D}_{n-1}^+ \cap \{x_1 \leq T - h\}$ via

$$\nabla_h^2 w(x, \mathbf{v}) = \frac{w(y_+) + w(\xi_{n-1}(y_-)) - h\mathbb{1}_{\xi_{n-1}}(y_-) - 2w(x)}{h^2}, \tag{3.11}$$

where $y_+ = \pi_{n-1}(x + h\mathbf{v})$ and $y_- = \pi_{n-1}(x - h\mathbf{v})$. By Lemma 3.3 the expression on the right-hand side of (3.11) involves evaluating w only at points in the sector $\mathbb{D}_{n-1}^+ \cap \{x_1 \leq T\}$. This is essentially equivalent to setting boundary conditions on the sector $\mathbb{D}_{n-1}^+$, though the explicit identification of those boundary conditions is a complicated task that we do not undertake here.

Now, the number of grid points in the full computational grid $[-T, T]^n \cap \mathbb{Z}_h^n$ grows exponentially in the dimension n as $\mathcal{O}((2Th^{-1})^n)$, which is known as the *curse of dimensionality*. However, the number of grid points in the sector $\mathbb{D}_{n-1}^+ \cap \{x_1 \leq T\}$ grows much slower in n, as the following lemma shows.

Lemma 3.4. *Let $G_{n,T}$ denote the number of grid points in the sector $\mathbb{D}_n^+ \cap \{x_1 \leq T\}$, where $T = mh$ with m a positive integer. Then $G_{n,T} \leq \frac{1}{n!}(Th^{-1} + n)^n$.*

https://doi.org/10.1017/S0956792525000075 Published online by Cambridge University Press

Proof. *Let $T = mh$ and define*

$$a_{m,n} = \sum_{i_1=0}^{m} \sum_{i_2=0}^{i_1} \cdots \sum_{i_{n-1}=0}^{i_{n-2}} \sum_{i_n=0}^{i_{n-1}} 1. \tag{3.12}$$

Note that $G_{n,T} = a_{m,n}$. Hence, the proof boils down to showing that $a_{m,n} \leq \frac{1}{n!}(m+n)^n$. We will prove this with induction on n, using the recursive identity

$$a_{m,n} = \sum_{i=0}^{m} a_{i,n-1},$$

which follows directly from (3.12). For the base step of $n = 1$, we have $a_{m,1} = m + 1$ by direct computation. For the inductive step, assume that for some $n \geq 1$ and all m we have $a_{m,n} \leq \frac{1}{n!}(m+n)^n$. Then we compute

$$a_{m,n+1} = \sum_{i=0}^{m} a_{i,n} \leq \sum_{i=0}^{m} \frac{1}{n!}(i+n)^n \leq \int_0^{m+1} \frac{1}{n!}(x+n)^n \, dx = \frac{1}{(n+1)!}(m+n+1)^{n+1},$$

which completes the inductive step and hence the proof. □

If we choose h small enough so that $n \leq Th^{-1}$, then the bound in Lemma 3.4 implies

$$G_{n,T} \leq \frac{(2Th^{-1})^n}{n!}.$$

This is a factor of $n!$ smaller than the number of grid points on the full computational domain $[-T, T]^n \cap \mathbb{Z}_h^n$, which reflects the exponentially small size of the sector $\mathbb{D}_n^+$. In order to understand better how $G_{n,T}$ scales with n in Lemma 3.4, we use a version of Stirling's formula $n! \geq \sqrt{2\pi n}(n/e)^n$ [41] to obtain

$$G_{n,T} \leq \frac{1}{n!}(Th^{-1} + n)^n \leq \frac{e^n}{\sqrt{2\pi n}}(1 + Th^{-1}n^{-1})^n \leq \frac{e^{n+Th^{-1}}}{\sqrt{2\pi n}}.$$

While this complexity is still exponential in n and h^{-1}, these two quantities do not directly interact. For example, with the full grid $[-T, T]^n \cap \mathbb{Z}_h^n$, increasing from n to $n+1$ dimensions requires a factor of $\mathcal{O}(h^{-1})$ more grid points, while for the sector $\mathbb{D}_n^+ \cap \{x_1 \leq T\}$ we require at most e times as many grid points, which is independent of the grid resolution h. However, it is important to point out that the curse of dimensionality is not overcome. It is rather more accurate to say that we have postponed the curse to larger values of n. For example, in Section 4, we conduct experiments with up to $n = 7$ experts, using the reduced computational grid in dimension $n - 1 = 6$, with a grid resolution of $h = 0.1$. Using the full computational grid we are only able to compute solutions of the $n = 4$ expert problem, on the reduced $n - 1 = 3$ dimensional grid.

Remark 3.5. *It is important to point out that the computational costs are more expensive than the memory required to store the grid, since we must compute the maximum over $\mathbf{v} \in \mathcal{B}^n$ in the operator, which is a maximum over 2^n directions. Thus, the computational cost admits an additional 2^n factor over the memory costs.*

4 Numerical results

In this section, we present numerical results for $n = 2$ up to $n = 10$ experts. In all cases, we solve the reduced equation (3.4) for $w(x) = u(x, 0)$, which is a problem in $n - 1$ dimensions. In Section 4.1 we present experiments on the full computational grid, while in Section 4.2 we present results on the smaller and more efficient sector grid. In each case, to solve the equation $w - F(\nabla_h^2 w) = g$, we iterate the fixed point scheme

$$w_{k+1} = (1 - dt)w_k + dt(F(\nabla_h^2 w_k) + g) \tag{4.1}$$

https://doi.org/10.1017/S0956792525000075 Published online by Cambridge University Press

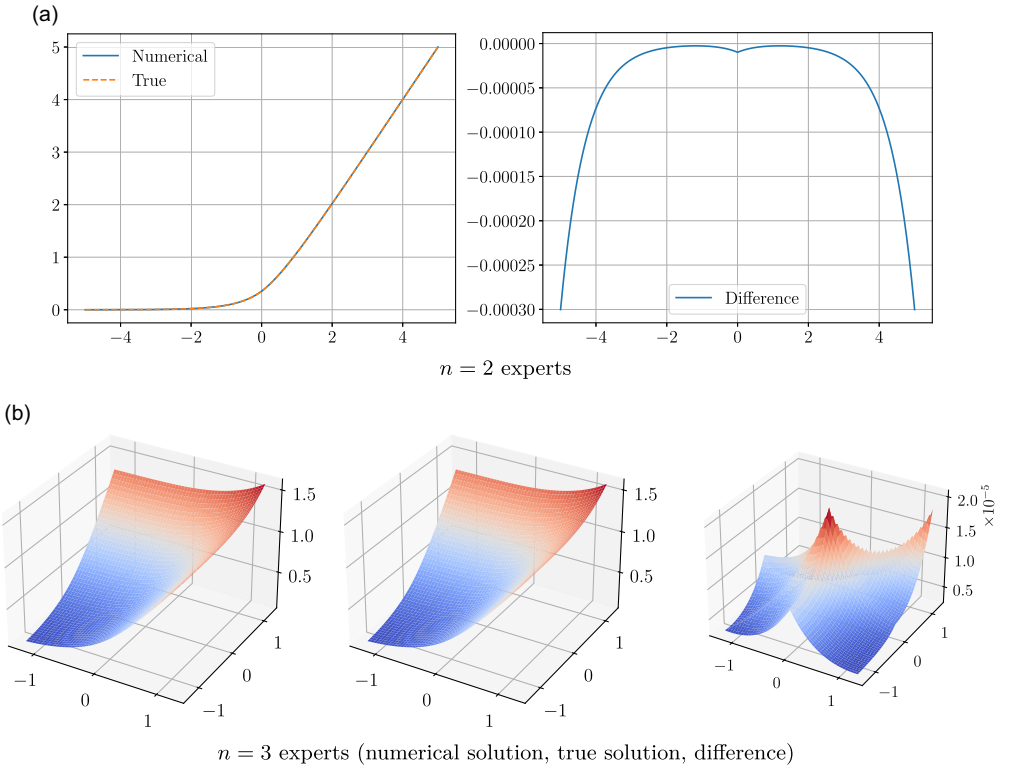

Figure 1. *Plots of the numerical solution w versus the true solutions for $n = 2$ and $n = 3$ experts.*

until

$$|w - F(\nabla_h^2 w) - g| \leq \frac{h^2}{100}.$$

In this section, we use F defined in (2.7). In all cases, we use $T = 5$ and restrict the solutions to the unit box $[0, 1]^{n-1}$. The code for all experiments is available on GitHub: https://github.com/jwcalder/PredictionPDE.

4.1 Full computational grid

We first present results on the full computational grid, so we solve the equation (3.6). Due to the curse of dimensionality, we can only solve the equation for $n = 2, 3, 4$ experts. The finest grid we used was $h = 0.01$ for $n \leq 3$ and $h = 0.025$ for $n = 4$. In Figure 1 we show plots of the numerical solutions versus the true solutions for $n = 2, 3$ experts. Since we solve the reduced $d = n - 1$ dimensional equation, these are PDEs in $d = 1$ and $d = 2$ dimensions. The $n = 2$ expert solution is accurate on the full domain $[-5, 5]$ since the boundary condition $u(\pm 5) = \max\{5, 0\}$ is exponentially accurate; see (1.6). For $n = 3$ experts, the solution loses accuracy away from the restricted domain $[-1, 1]^2$.

In Figure 2 (a) we show a convergence analysis for varying grid resolution h for $n = 2, 3, 4$ experts, where the exact solutions of the PDEs are known. In all cases, we observe second-order $\mathcal{O}(h^2)$ convergence rates. Figure 2 also shows the optimality of each adversarial strategy. We define the optimality of strategy $\mathbf{v} \in \mathcal{B}^{n-1}$ at a grid point x by the ratio

$$\text{opt}(x, \mathbf{v}) = \frac{\nabla_h^2 w(x, \mathbf{v})}{\max_{\mathbf{p} \in \mathcal{B}^{n-1}} \nabla_h^2 w(x, \mathbf{p})} = \frac{\nabla_h^2 w(x, \mathbf{v})}{2(u(x) - g(x, 0))}.$$

https://doi.org/10.1017/S0956792525000075 Published online by Cambridge University Press

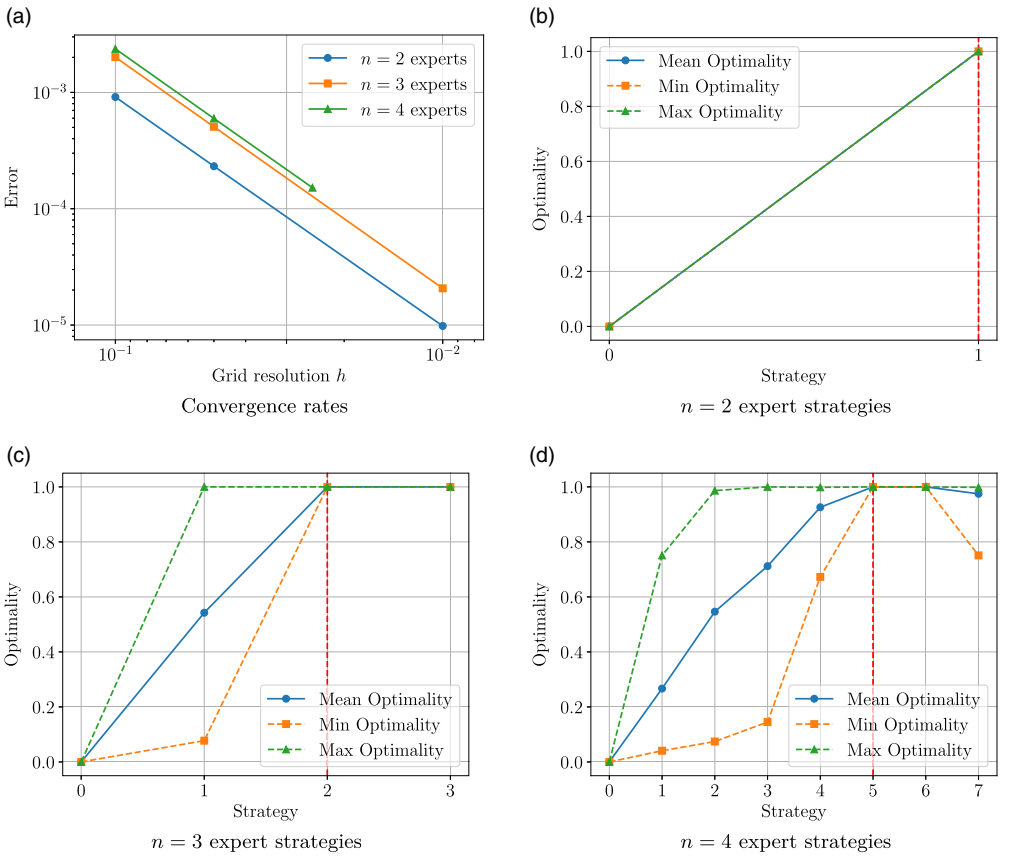

Figure 2. *Convergence rates and optimal strategies for $n = 2, 3, 4$ experts, computed from the numerical solutions. The dashed red line indicates the COMB strategy, which is numerically observed to be optimal for $n = 2, 3, 4$ experts, as the theory predicts.*

An optimality value of $\text{opt}(x, \mathbf{v}) = 1$ indicates that strategy $\mathbf{v}$ is optimal at grid point x. In order to measure the optimality of all the competing strategies, we plot the minimum, maximum, and average optimality values over the intersection of the computational grid with the positive sector $\mathbb{D}^+_{n-1}$. Since the solution is permutation invariant, the scores are the same in all other sectors. If the *minimum* score is 1, up to the numerical precision $O(h^2)$, then that strategy is globally optimal over the unit box $[-1, 1]^{n-1}$. In Figure 2 we denote the strategies, which are binary vectors, by the decimal number that strategy corresponds to, and we indicate the COMB strategy with a dashed red line. Since $\mathbf{v}$ and $\mathbb{1} - \mathbf{v}$ are equivalent strategies, we only show the optimality scores for the first half of the strategies; the plot for the second half is a mirror image of the plots shown.

In Figure 2 (b), we see that strategies $1 = (0, 1)$ and $2 = (1, 0)$ are optimal, both of which correspond to the COMB strategy. In Figure 2 (c), we see that for $n = 3$ experts, the COMB strategy $2 = (0, 1, 0)$ is optimal, as well as the non-COMB strategy $3 = (0, 1, 1)$. In this case, we recall that the complement strategies $(1, 0, 1)$ and $(1, 0, 0)$ are equivalent, and hence also optimal, but are not depicted. For $n = 4$ experts in Figure 2 (d), we see that the COMB strategy $5 = (0, 1, 0, 1)$ is optimal, as well as the non-COMB strategy $6 = (0, 1, 1, 0)$. All of these results have already been established theoretically in previous work; see Section 1.1. We presented these results to verify that the numerical solvers are working properly and give results that agree with previous work.

We are unable to solve the PDE (3.6) for w on a full computational grid for the $n = 5$ expert problem, so for this we must resort to the sparse grid method that restricts attention to the positive sector.

https://doi.org/10.1017/S0956792525000075 Published online by Cambridge University Press

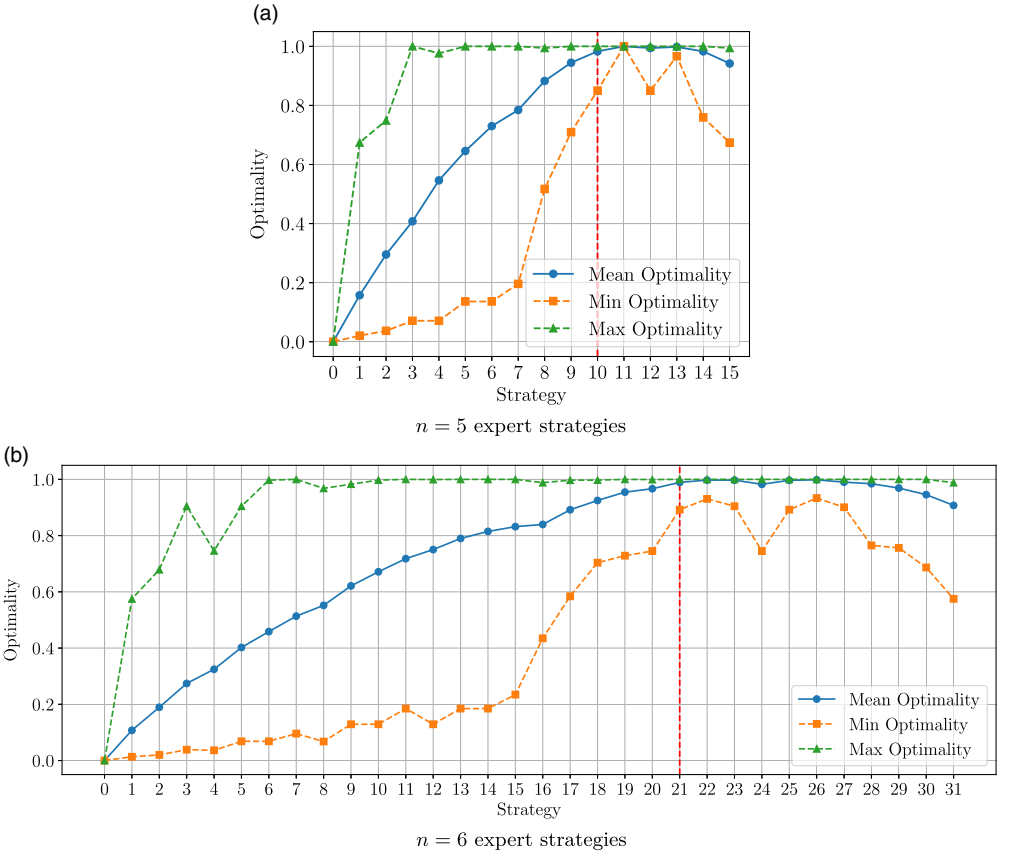

Figure 3. *Numerical computation of strategy optimality for the $n = 5, 6$ expert problems.*

4.2 Sparse grids

We now turn to computations on the positive sector $\mathbb{D}^+_{n-1} \cap [0, T]^{n-1}$, which is far smaller than the full grid and allows us to run experiments for $n = 5, 6, 7, 8, 9, 10$ experts. In this case, we are solving equation (3.10) using the methods described in Section 3.2. To facilitate computations, we flatten the sector $\mathbb{D}^+_{n-1} \cap [0, T]^{n-1}$ to a one-dimensional array, and store the solution w as a one-dimensional array with linear indexing. We pre-computed and stored the stencils for second derivatives in all directions $\mathbf{v} \in \mathcal{B}^{n-1}$ using the methods outlined in Section 3.2 prior to the running the iteration (4.1) to solve the equation. All code is written in Python using the Numpy package and fully vectorized operations.

We used grid resolutions of $h = 0.025$ for $n = 5$, $h = 0.05$ for $n = 6$ and $h = 0.1$ for $n = 7$, $h = 0.2$ for $n = 8$, $h = 0.25$ for $n = 9$, and $h = 0.35$ for $n = 10$ experts. Table 1 shows the number of grid points used in each dimension, compared to the number of grid points that would be required on the full grid. The simulations required between 25 GB and 75 GB of memory and each took less than one day to run on a single processor. Figures 3, 4, and 5 show the numerically computed optimality of the adversary's strategies for $n = 5, 6, 7, 8, 9, 10$ experts. In all cases, we have strong numerical evidence to indicate that the COMB strategy is *not* globally optimal. This corroborates numerical evidence from [17] for $n = 5$.

We have strong numerical evidence in Figure 3 that the strategy $11 = (0, 1, 0, 1, 1)$ is optimal for the $n = 5$ expert problem over the box $[-1, 1]^5$. Furthermore, the numerical evidence points to this being the *only* optimal strategy, over the unit box, for $n = 5$ experts. The minimum optimality score for strategy 11 is 0.9999999999991459, which is far more accurate than the second-order $O(h^2)$ accuracy for $h = 0.05$ would suggest. The second best competing strategy is $13 = (0, 1, 1, 0, 1)$ with a minimum optimality

https://doi.org/10.1017/S0956792525000075 Published online by Cambridge University Press

Table 1. *Number of grid points in the sector computational domain $\mathbb{D}_d^+ \cap [0,5]^d$ compared to the full grid $\mathbb{Z}_h^d \cap [-5,5]^d$. We use fewer grid points as the dimension increases since evaluating the PDE involves computing derivatives in 2^d directions, so the computational time and memory storage increase exponentially with d*

Dimension ($d = n-1$)	Grid resolution h	Sector $\mathbb{D}_d^+ \cap [0,5]^d$	Full grid $\mathbb{Z}_h^d \cap [-5,5]^d$
4	0.025	7×10^7	3×10^{10}
5	0.050	1×10^8	3×10^{11}
6	0.100	3×10^7	1×10^{12}
7	0.200	3×10^6	8×10^{11}
8	0.250	3×10^6	7×10^{12}
9	0.350	8×10^5	1×10^{13}

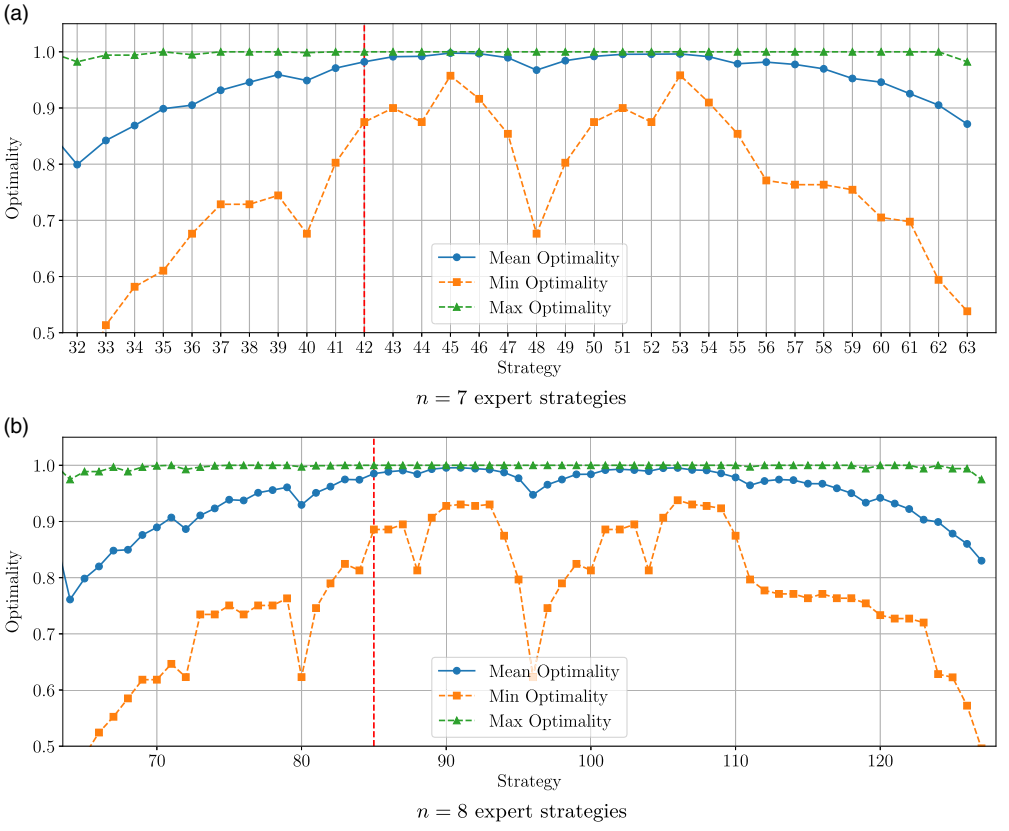

Figure 4. *Numerical computation of strategy optimality for the $n = 7, 8$ expert problems.*

score of 0.966, which is well outside the range of numerical precision, suggesting that strategy 13 is not globally optimal.

The remaining plots in Figures 3, 4, and 5 provide very strong numerical evidence that there are no globally optimal adversarial strategies for $n = 6, 7, 8, 9, 10$ experts. The highest minimum optimality scores are well outside of numerical precision. The one exception is the $n = 10$ expert problem, where there are strategies with minimum optimality scores above 0.98, while the grid resolution of $h = 0.35$ yields $h^2 = 0.1225$. However, we expect this is an artefact from using a very coarse grid. We observed a similar phenomenon with $n = 9$ experts, where some strategies appeared more optimal on a coarse grid.

https://doi.org/10.1017/S0956792525000075 Published online by Cambridge University Press

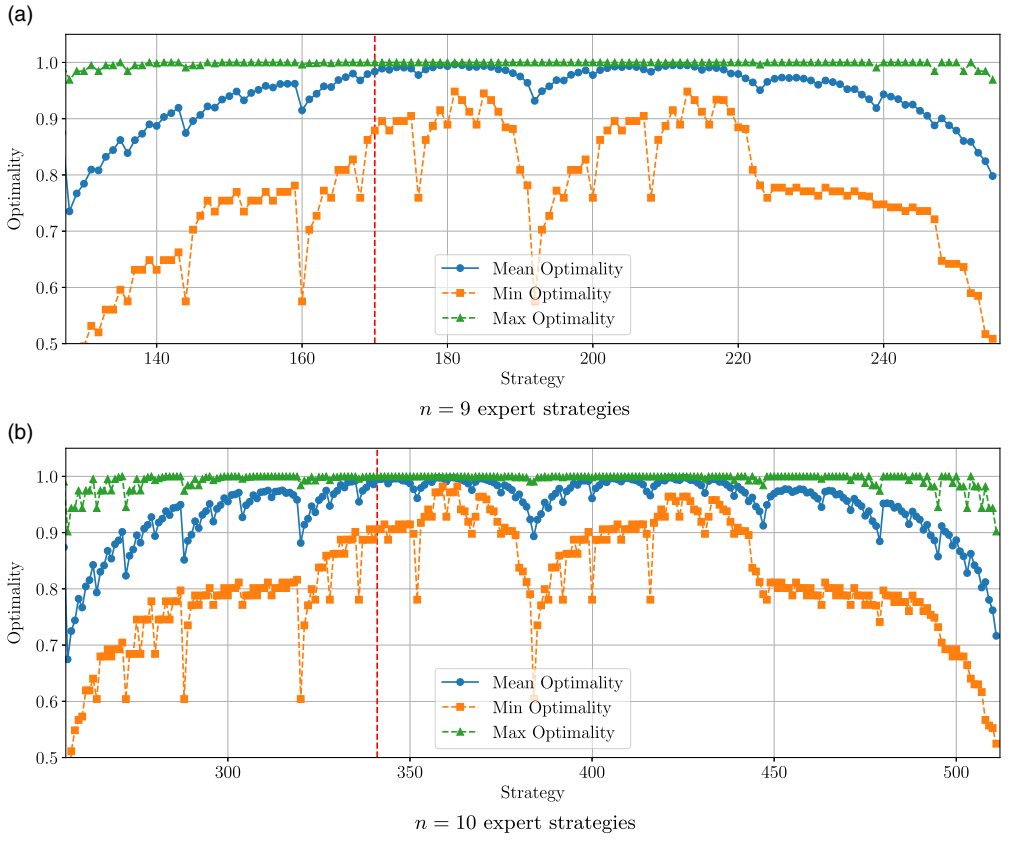

Figure 5. *Numerical computation of strategy optimality for the $n = 9, 10$ expert problems.*

5 Conclusion and future work

This paper developed and analysed a numerical scheme to solve a degenerate elliptic PDE arising from prediction with expert advice in relatively high dimensions ($n \leq 10$) by exploiting symmetries in the equation and solution. Based on numerical results, we are able to make a number of conjectures for the optimality of various adversarial strategies in Section 1.2. Our results have some limitations; mainly we are not able to solve the PDE on all of $\mathbb{R}^n$, and so our results are restricted to the box $[-1, 1]^n$.

We expect these numerical methods could be extended to a few more experts, perhaps the $n = 11$ and $n = 12$ expert problems, using parallel processing or computational clusters with vastly more memory. The finite horizon problem should be amenable to similar techniques. There are also other prediction with expert advice PDEs, in particular the history-dependent experts setting [12, 21], that would benefit from numerical explorations. In terms of theory, we posed a number of conjectures and open problems that stem from this work in Section 1.2 that would be interesting to explore in future work.

Funding. Calder and Mosaphir were partially supported by NSF-DMS grant 1944925, and Calder was partially supported by the Alfred P. Sloan Foundation, a McKnight Presidential Fellowship, and the Albert and Dorothy Marden Professorship. Drenska was partially supported by NSF-DMS grant 2407839.

Data availability statement. https://github.com/jwcalder/PredictionPDE

Competing interests. The authors declare that they have no competing interests.

https://doi.org/10.1017/S0956792525000075 Published online by Cambridge University Press

References

[1] Abbasi-Yadkori, Y., Bartlett, P. L. & Gabillon, V. (2017) Near minimax optimal players for the finite-time 3-expert prediction problem. In Advances in Neural Information Processing Systems, pp. 3033–3042.

[2] Amin, K., Kale, S., Tesauro, G. & Turaga, D. (2015) Budgeted prediction with expert advice. In Twenty-Ninth AAAI Conference on Artificial Intelligence.

[3] Antunovic, T., Peres, Y., Sheffield, S. & Somersille, S. (2012) Tug-of-war and infinity Laplace equation with vanishing Neumann boundary condition. *Commun. Partial Differ. Equ.* **37**(10), 1839–1869.

[4] Armstrong, S. N. & Smart, C. K. (2012) A finite difference approach to the infinity Laplace equation and tug-of-war games. *Trans. Amer. Math. Soc.* **364**(2), 595–636.

[5] Bayraktar, E., Ekren, I. & Zhang, X. (2020) Finite-time 4-expert prediction problem. *Commun. Partial Differ. Equ.* **45**(7), 714–757.

[6] Bayraktar, E., Ekren, I. & Zhang, X. (2021) Prediction against a limited adversary. *J. Mach. Learn. Res.* **22**(72), 1–33.

[7] Bayraktar, E., Ekren, I. & Zhang, Y. (2020) On the asymptotic optimality of the comb strategy for prediction with expert advice. *Ann. Appl. Probab.* **30**(6), 2517–2546.

[8] Bayraktar, E., Poor, H. V. & Zhang, X. (2020) Malicious experts versus the multiplicative weights algorithm in online prediction. *IEEE Trans. Inf. Theory* **67**(1), 559–565.

[9] Calder, J. (2018) The game theoretic p-Laplacian and semi-supervised learning with few labels. *Nonlinearity* **32**(1), 301–330.

[10] Calder, J. (2018) Lecture notes on viscosity solutions. Online Lecture Notes, http://www-users.math.umn.edu/

[11] Calder, J. (2019) Consistency of Lipschitz learning with infinite unlabeled data and finite labeled data. *SIAM J. Math. Data Sci.* **1**(4), 780–812.

[12] Calder, J. & Drenska, N. (2021) Asymptotically optimal strategies for online prediction with history-dependent experts. *J. Fourier Anal. Appl. Special Collection on Harmonic Anal. Comb. Graphs* **27**(20).

[13] Calder, J. & Drenska, N. (2024). Consistency of semi-supervised learning, stochastic tug-of-war games, and the p-Laplacian. In: Carrillo, J.A., Tadmor, E. (eds) Active Particles, Volume 4. Modeling and Simulation in Science, Engineering and Technology. Birkhäuser, Cham.

[14] Calder, J. & Smart, C. K. (2020) The limit shape of convex hull peeling. *Duke Math. J.* **169**(11), 2079–2124.

[15] Cesa-Bianchi, N., Freund, Y., Haussler, D., Helmbold, D. P., Schapire, R. E. & Warmuth, M. K. (1997) How to use expert advice. *J. ACM* **44**(3), 427–485. May

[16] Cesa-Bianchi, N. & Lugosi, G. (2006) *Prediction, Learning, and Games*, Cambridge University Press, New York, NY, USA.

[17] Chase, Z. (2019) Experimental evidence for asymptotic non-optimality of comb adversary strategy. arXiv preprint arXiv: 1912.01548

[18] Cover, T. M. (1966) Behavior of sequential predictors of binary sequences. Technical report, Stanford University California Stanford Electronics Labs

[19] Crandall, M. G., Ishii, H. & Lions, P.-L. (1992) User's guide to viscosity solutions of second order partial differential equations. *Bull. Am. Math. Soc.* **27**(1), 1–67.

[20] Drenska, N. (2017) A PDE approach to a prediction problem involving randomized strategies, *PhD thesis*, New York University

[21] Drenska, N. & Calder, J. (2023) Online prediction with history-dependent experts: The general case. *Commun. Pure Appl. Math.* **76**(9), 1678–1727.

[22] Drenska, N. & Kohn, R. V. (2020) Prediction with expert advice: A PDE perspective. *J. Nonlinear Sci.* **30**(1), 137–173.

[23] Drenska, N. & Kohn, R. V. (2023) A pde approach to the prediction of a binary sequence with advice from two history-dependent experts. *Commun. Pure Appl. Math.* **76**(4), 843–897.

[24] Flores, M., Calder, J. & Lerman, G. (2022) Analysis and algorithms for lp-based semi-supervised learning on graphs. *Appl. Comput. Harmon. Anal.* **60**, 77–122.

[25] Freund, Y. & Schapire, R. E. (1997) A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Syst. Sci.* **55**(1), 119–139.

[26] Friedman, A. (2013). *Differential Games*. Courier Corporation, Dover Publications, Mineola New York.

[27] Gravin, N., Peres, Y. & Sivan, B. (2016) Towards optimal algorithms for prediction with expert advice. In Proceedings of the twenty-seventh annual ACM-SIAM symposium on Discrete algorithms, pp. 528–547. SIAM.

[28] Gravin, N., Peres, Y. & Sivan, B. (2017) Tight lower bounds for multiplicative weights algorithmic families. In 44th International Colloquium on Automata, Languages, and Programming (ICALP 2017). Schloss Dagstuhl-Leibniz-Zentrum Fuer Informatik.

[29] Hannan, J. (1957) Approximation to bayes risk in repeated play. *Contrib. Theory Games* **3**, 97–139.

[30] Haussler, D., Kivinen, J. & Warmuth, M. K. (1995) Tight worst-case loss bounds for predicting with expert advice. In European Conference on Computational Learning Theory, pp. 69–83. Springer.

[31] Kobzar, V. A., Kohn, R. V. & Wang, Z. (2020) New potential-based bounds for prediction with expert advice. In Conference on Learning Theory, pp. 2370–2405.PMLR.

[32] Kobzar, V. A., Kohn, R. V. & Wang, Z. (2020) New potential-based bounds for the geometric-stopping version of prediction with expert advice. In Mathematical and Scientific Machine Learning, pp. 537–554. PMLR.

[33] Kohn, R. V. & Serfaty, S. (2006) A deterministic-control-based approach motion by curvature. *Commun Pur Appl Math* **59**(3), 344–407.

[34] Kohn, R. V. & Serfaty, S. (2010) A deterministic-control-based approach to fully nonlinear parabolic and elliptic equations. *Commun. Pure Appl. Math.* **63**(10), 1298–1350.

https://doi.org/10.1017/S0956792525000075 Published online by Cambridge University Press

[35] Lewicka, M. & Manfredi, J. J. (2015) The obstacle problem for the p-Laplacian via optimal stopping of tug-of-war games. *Probab. Theory Relat. Fields* **167**, 349–378.
[36] Littlestone, N. & Warmuth, M. K. (1994) The weighted majority algorithm. *Inf. Comput.* **108**(2), 212–261. Feb
[37] Naor, A. & Sheffield, S. (2012) Absolutely minimal lipschitz extension of tree-valued mappings. *Math. Ann.* **354**(3), 1049–1078.
[38] Oberman, A. M. (2006). Convergent difference schemes for degenerate elliptic and parabolic equations: Hamilton–Jacobi equations and free boundary problems. *SIAM J. Numer. Anal.* **44**(2), 879–895.
[39] Peres, Y., Schramm, O., Sheffield, S. & Wilson, D. B. (2009) Tug-of-war and the infinity Laplacian. *J. Amer. Math. Soc* **22**(1), 167–210.
[40] Peres, Y. & Sheffield, S. (2008) Tug-of-war with noise: A game-theoretic view of the p -laplacian. *Duke Math. J.* **145**(1), 91–120. 10
[41] Robbins, H. (1955) A remark on Stirling's formula. *Amer. Math. Monthly* **62**(1), 26–29.
[42] Rokhlin, D. (2017) PDE approach to the problem of online prediction with expert advice: A construction of potential-based strategies. *Int. J. Pure Appl. Math.* **114**(4), 05.
[43] Wang, Z. (2021). PDE approaches to two online learning problems, and an empirical study of some neural network-based active learning algorithms, *PhD thesis*, New York University

A Definition of a viscosity solution

For convenience, we recall the definitions of viscosity solutions for a general second-order nonlinear partial differential equation

$$H(\nabla^2 u, \nabla u, u, x) = 0 \quad \text{in } \mathcal{O}, \tag{A.1}$$

where H is continuous and $\mathcal{O} \subset \mathbb{R}^n$. Let USC($\mathcal{O}$) (resp. LSC($\mathcal{O}$)) denote the collection of functions that are upper (resp. lower) semicontinuous at all points in $\mathcal{O}$. We make the following definitions.

We first recall the test function definition of viscosity solution.

Definition A.1 (Viscosity solution). *We say that $u \in USC(\mathcal{O})$ is a viscosity subsolution of (A.1) if for every $x \in \mathcal{O}$ and every $\varphi \in C^\infty(\mathbb{R}^n)$ such that $u - \varphi$ has a local maximum at x with respect to $\mathcal{O}$*

$$H(\nabla^2\varphi(x), \nabla\varphi(x), u(x), x) \leq 0.$$

We will often say that $u \in USC(\mathcal{O})$ is a viscosity solution of $H \leq 0$ in $\mathcal{O}$ when u is a viscosity subsolution of (A.1).

Similarly, we say that $u \in LSC(\mathcal{O})$ is a viscosity supersolution of (A.1) if for every $x \in \mathcal{O}$ and every $\varphi \in C^\infty(\mathbb{R}^n)$ such that $u - \varphi$ has a local minimum at x with respect to $\mathcal{O}$

$$H(\nabla^2\varphi(x), \nabla\varphi(x), u(x), x) \geq 0.$$

We also say that $u \in LSC(\mathcal{O})$ is a viscosity solution of $H \geq 0$ in $\mathcal{O}$ when u is a viscosity supersolution of (A.1).

Finally, we say u is viscosity solution of (A.1) if u is both a viscosity subsolution and a viscosity supersolution.

For more details on the rich theory of viscosity solutions, we refer the reader to the user's guide [19] and [10].

Cite this article: Calder J., Drenska N. and Mosaphir D. Numerical solution of a PDE arising from prediction with expert advice. *European Journal of Applied Mathematics*, https://doi.org/10.1017/S0956792525000075