Convergence Analysis of Prediction Markets via Randomized Subspace Descent

Rafael Frongillo
Department of Computer Science
University of Colorado, Boulder
`raf@colorado.edu`

Mark D. Reid
Research School of Computer Science
The Australian National University & NICTA
`mark.reid@anu.edu.au`

Abstract

Prediction markets are economic mechanisms for aggregating information about future events through sequential interactions with traders. The pricing mechanisms in these markets are known to be related to optimization algorithms in machine learning and through these connections we have some understanding of how equilibrium market prices relate to the beliefs of the traders in a market. However, little is known about rates and guarantees for the convergence of these sequential mechanisms, and two recent papers cite this as an important open question.

In this paper we show how some previously studied prediction market trading models can be understood as a natural generalization of randomized coordinate descent which we call *randomized subspace descent* (RSD). We establish convergence rates for RSD and leverage them to prove rates for the two prediction market models above, answering the open questions. Our results extend beyond standard centralized markets to arbitrary trade networks.

1 Introduction

In recent years, there has been an increasing appreciation of the shared mathematical foundations between prediction markets and a variety of techniques in machine learning. Prediction markets consist of agents who trade securities that pay out depending on the outcome of some uncertain, future event. As trading takes place, the prices of these securities reflect an aggregation of the beliefs the traders have about the future event. A popular class of mechanisms for updating these prices as trading occurs has been shown to be closely related to techniques from online learning [7, 1, 21], convex optimization [10, 19, 13], probabilistic aggregation [24, 14], and crowdsourcing [3]. Building these connections serve several purposes, however one important line of research has been to use insights from machine learning to better understand how to interpret prices in a prediction market as aggregations of trader beliefs, and moreover, how the market together with the traders can be viewed as something akin to a distributed machine learning algorithm [24].

The analysis in this paper was motivated in part by two pieces of work that considered the equilibria of prediction markets with specific models of trader behavior: traders as risk minimizers [13]; and traders who maximize expected exponential utility using beliefs from exponential families [2]. In both cases, the focus was on understanding the properties of the market at convergence, and questions concerning whether and how convergence happened were left as future work. In [2], the authors note that "we have not considered the dynamics by which such an equilibrium would be reached, nor the rate of convergence etc., yet we think such questions provide fruitful directions for future research." In [13], "One area of future work would be conducting a detailed analysis of this framework using the tools of convex optimisation. A particularly interesting topic is to find the conditions under which the market will converge."

The main contribution of this paper is to answer these questions of convergence. We do so by first proposing a new and very general model of trading networks and dynamics (§3) that subsumes the models used in [2] and [13] and provide a key structural result for what we call *efficient* trades in these networks (Theorem 2). As an aside, this structural result provides an immediate generalization of an existing aggregation result in [2] to trade networks of "compatible" agents (Theorem 8). In §4, we argue that efficient trades in our networks model can be viewed as steps of what we call *Random Subspace Descent* (RSD) algorithm (Algorithm 1). This novel generalization of coordinate descent allows an objective to be minimized by taking steps along affinely constrained subspaces, and maybe be of independent interest beyond prediction market analysis. We provide a convergence analysis of RSD under two sets of regularity constraints (Theorems 3 & 9) and show how these can be used to derive (slow & fast) convergence rates in trade networks (Theorems 4 & 5).

Before introducing our general trading networks and convergence rate results, we first introduce the now standard presentation of *potential-based prediction markets* [1] and the recent variant in which all agents determine their trades using *risk measures* [13]. We will then state informal versions of our main results so as to highlight how we address issues of convergence in existing frameworks.

2 Background and Informal Results

Prediction markets are mechanisms for eliciting and aggregating distributed information or beliefs about uncertain future events. The set of events or *outcomes* under consideration in the market will be denoted Ω and may be finite or infinite. For example, each outcome $\omega \in \Omega$ might represent a certain presidential candidate winning an election, the location of a missing submarine, or an unknown label for an item in a data set. Following [1], the goods that are traded in a prediction market are k outcome-dependent *securities* $\{\phi(\cdot)_i\}_{i=1}^k$ that pay $\phi(\omega)_i$ dollars should the outcome $\omega \in \Omega$ occur. We denote the set of distributions over Ω by Δ_Ω and note, for any $p \in \Delta_\Omega$, that the expected pay off for the securities under p is $\mathbb{E}_{\omega \sim p}[\phi(\omega)]$ and the set of all expected pay offs is just the convex hull, denoted $\Pi := \operatorname{conv}(\phi(\Omega))$. A simple and commonly studied case is when $\Omega = [k] := \{1, \ldots, k\}$ (*i.e.*, when there are exactly k outcomes) and the securities are the *Arrow-Debreu* securities that pay out $1 should a specific outcome occur and nothing otherwise (*i.e.*, $\phi(\omega)_i = 1$ if $\omega = i$ and $\phi(\omega)_i = 0$ for $\omega \neq i$). Here, the securities are just basis vectors for $\mathbb{R}^k$ and $\Pi = \Delta_\Omega$.

Traders in a prediction market hold portfolios of securities $r \in \mathbb{R}^k$ called *positions* that pay out a total of $r \cdot \phi(\omega) = \sum_{i=1}^k r_i \phi(\omega)_i$ dollars should outcome ω occur. We denote the set of positions by $\mathcal{R} = \mathbb{R}^k$. We will assume that $\mathcal{R}$ always contains a position $r^\$$ that returns a dollar regardless of which outcome occurs, meaning $r^\$ \cdot \phi(\omega) = 1$ for all $\omega \in \Omega$. We therefore interpret $r^\$$ as "cash" within the market in the sense that buying or selling $r^\$$ guarantees a fixed change in wealth.

In order to address the questions about convergence in [2, 13] we will consider a common form of prediction market that is run through a *market maker*. This is an automated agent that is willing to buy or sell securities in return for cash. The specific and well-studied prediction market mechanism we consider is the *potential-based market maker* [1]. Here, traders interact with the market maker sequentially, and the cost for each trade is determined by a convex potential function $C : \mathcal{R} \to \mathbb{R}$ applied to the market maker's *state* $s \in \mathcal{R}$. Specifically, the cost for a trade dr when the market maker has state s is given by $\operatorname{cost}(dr; s) = C(s - dr) - C(s)$, *i.e.*, the change in potential value of the market maker's position due to the market maker accepting the trade. After a trade, the market maker updates the state to $s \leftarrow s - dr$.[1] As noted in the next section, the usual axiomatic requirements for a cost function (*e.g.*, in [1]) specify a function that is effectively a *risk measure*, commonly studied in mathematical finance (see, *e.g.*, [9]).

2.1 Risk Measures

As in [13], agents in our framework will each quantify their uncertainty in positions using what is known as risk measure. This is a function that assigns dollar values to positions. As Example 1 below shows, this assumption will also cover the case of agents maximizing exponential utility, as considered in [2].

---

[1]It is more common in the prediction market literature for s to be a *liability* vector, tracking what the market maker stands to lose instead of gain. Here we adopt positive positions to match the convention for risk measures.

A (convex monetary) *risk measure* is a function $\rho : \mathcal{R} \to \mathbb{R}$ satisfying, for all $r, r' \in \mathcal{R}$:

- *Monotonicity*: $\forall \omega \ \ r \cdot \phi(\omega) \leq r' \cdot \phi(\omega) \implies \rho(r) \geq \rho(r')$.
- *Cash invariance*: $\rho(r + c\,r^\$) = \rho(r) - c$ for all $c \in \mathbb{R}$.
- *Convexity*: $\rho\big(\lambda r + (1 - \lambda)r'\big) \leq \lambda\rho(r) + (1 - \lambda)\rho(r')$ for all $\lambda \in (0, 1)$.
- *Normalization*: $\rho(0) = 0$.

The reasonableness of these properties is usually argued as follows (see, *e.g.*, [9]). Monotonicity ensures that positions that result in strictly smaller payoffs regardless of the outcome are considered more risky. Cash invariance captures the idea that if a guaranteed payment of \$$c$ is added to the payment on each outcome then the risk will decrease by \$$c$. Convexity states that merging positions results in lower risk. Finally, normalization requires that holding no securities should carry no risk. This last condition is only for convenience since any risk without this condition can trivially have its argument translated so it holds without affecting the other three properties. A key result concerning convex risk measures is the following representation theorem (cf. [9, Theorem 4.15],).

Theorem 1 (Risk Representation). *A functional $\rho : \mathcal{R} \to \mathbb{R}$ is a convex risk measure if and only if there is a closed convex function $\alpha : \Pi \to \mathbb{R} \cup \{\infty\}$ such that $\rho(r) = \sup_{\pi \in \mathrm{relint}(\Pi)} \langle \pi, -r \rangle - \alpha(\pi)$.*

Here $\mathrm{relint}(\Pi)$ denotes the relative interior of Π, the interior relative to the affine hull of Π. Notice that if f^* denotes the convex conjugate $f^*(y) := \sup_x \langle y, x \rangle - f(x)$, then this theorem states that $\rho(r) = \alpha^*(-r)$, that is, ρ and α are "dual" in the same way prices and positions are dual [5, §5.4.4]. This suggests that the function α can be interpreted as a *penalty function*, assigning a measure of "unlikeliness" $\alpha(\pi)$ to each expected value π of the securities defined above. Equivalently, $\alpha(\mathbb{E}_p[\phi])$ measures the unlikeliness of distribution p over the outcomes. We can then see that the risk is the greatest expected loss under each distribution, taking into account the penalties assigned by α.

Example 1. *A well-studied risk measure is the* entropic risk *relative to a reference distribution $q \in \Delta_\Omega$ [9]. This is defined on positions $r \in \mathcal{R}$ by $\rho^\beta(r) := \beta \log \mathbb{E}_{\omega \sim q} [\exp(-r \cdot \phi(\omega)/\beta)]$. The cost function $C(r) = \rho^\beta(-r)$ associated with this risk exactly corresponds to the* logarithmic market scoring rule *(LMSR). Its associated convex function α^β over distributions is the scaled relative entropy $\alpha^\beta(p) = \beta \mathrm{KL}(p \,|\, q)$. As discussed in [2, 13], the entropic risk is closely related to* exponential utility $U_\beta(w) := -\frac{1}{\beta} \exp(-\beta w)$. *Indeed, $\rho^\beta(r) = -U_\beta \big(\mathbb{E}_{\omega \sim q} [U_\beta(r \cdot \phi(\omega))]\big)$ which is just the negative* certainty equivalent *of the position r — i.e., the amount of cash an agent with utility U_β and belief q would be willing to trade for the uncertain position r. Due to the monotonicity of U_β^{-1}, it follows that a trader maximizing expected utility $\mathbb{E}_{\omega \sim q} [U_\beta(r \cdot \phi(\omega))]$ of holding position r is equivalent to minimizing the entropic risk $\rho^\beta(r)$.*

For technical reasons, in addition to the standard assumptions for convex risk measures, we will also make two weak regularity assumptions. These are similar to properties required of cost functions in the prediction market literature (cf. [1, Theorem 3.2]):

- *Expressiveness*: ρ is everywhere differentiable, and $\mathrm{closure}\{\nabla\rho(r) : r \in \mathcal{R}\} = \Pi$.
- *Strict risk aversion*: the *Convexity* inequality is strict unless $r - r' = c\,r^\$$ for some $c \in \mathbb{R}$.

As discussed in [1], expressiveness is related to the dual formulation given above; roughly, it says that the agent must take into account every possible expected value of the securities when calculating the risk. Strict risk aversion says that an agent should strictly prefer a mixture of positions, unless of course the difference is outcome-independent.

Under these assumptions, the representation result of Theorem 1 and a similar result for cost functions [1, Theorem 3.2]) coincide and we are able to show that cost functions and risk measures are exactly the same object; we write $\rho_C(r) = C(r)$ when we think of C as a risk measure. Unfolding the definition of cost now using cash invariance, we have $\rho_C(s - dr + \mathrm{cost}(dr; s)r^\$) = \rho_C(s - dr) - \mathrm{cost}(dr; s) = C(s - dr) - C(s - dr) + C(s) = \rho_C(s)$. Thus, we may view a potential-based market maker as a *constant-risk* agent.

2.2 Trading Dynamics and Aggregation

As described above, we consider traders who approach the market maker sequentially and at random, and select the optimal trade based on their current position, the market state, and the cost function C.

As we just observed, we may think of the market maker as a constant-risk agent with $\rho_C = C$. Let us examine the optimization problem faced by the trader with position r when the current market state is s. This trader will choose a portfolio dr^* from the market maker so as to minimise her risk:

$$dr^* \in \underset{dr \in \mathbb{R}^k}{\arg\min} \, \rho \left(r + dr - \text{cost}(dr)r^\$ \right) = \underset{dr \in \mathbb{R}^k}{\arg\min} \, \rho(r + dr) + \rho_C(s - dr) . \tag{1}$$

Since, by the cash invariance of ρ and the definition of cost, the objective is $\rho(r + dr) + \rho_C(s - dr) - \rho_C(s)$, and $\rho_C(s)$ does not depend on dr. Thus, if we think of $F(r, s) = \rho(r) + \rho_C(s)$ as a kind of "social risk", we can define the *surplus* as simply the net risk taken away by an optimal trade, namely $F(r, s) - F(r + dr^*, s - dr^*)$.

We can now state our central question: *if a set of N such traders arrive at random and execute optimal (or perhaps near-optimal) trades with the market maker, will the market state converge to the optimal risk, and if so how fast?* As discussed in the introduction, this is precisely the question asked in [2, 13] that we set out to answer. To do so we will draw a close connection to the literature on distributed optimization algorithms for machine learning. Specifically, if we encode the entire state of our system in the positions $R = (r_0 = s, r_1, \ldots, r_n)$ of the market maker and each of the n traders, we may view the optimal trade in eq. (1) as performing a *coordinate descent* step, by optimizing only with respect to coordinates 0 and i. We build on this connection in Section 4 and leverage a generalization of coordinate descent methods to show the following in Theorem 4: *If a set of risk-based traders is sampled at random to sequentially trade in the market, the market state and prices converge to within ϵ of the optimal total risk in $O(1/\epsilon)$ rounds.*

In fact, under mild smoothness assumptions on the cost potential function C, we can improve this rate to $O(\log(1/\epsilon))$. We can also relax the optimality of the trader behavior; as long as traders find a trade dr which extracts at least a constant fraction of the surplus, the rate remains intact.

With convergence rates in hand, the next natural question might be: *to what* does the market converge? Abernethy et al. [2] show that when traders minimize expected exponential utility and have exponential family beliefs, the market equilibrium price can be thought of as a weighted average of the parameters of the traders, with the weights being a measure of their risk tolerance. Even though our setting is far more general than exponential utility and exponential families, the framework we develop can also be used to show that their results can be extended to interactions between traders who have what we call "compatible" risks and beliefs. Specifically, for any risk-based trader possessing a risk ρ with dual α, we can think of that trader's "belief" as the least surprising distribution p according to α. This view induces a family of distributions (which happen to be *generalized exponential families* [11]) that are parameterized by the initial positions of the traders. Furthermore, the risk tolerance b is given by how sensitive this belief is to small changes of an agent's position. The results of [2] are then a special case of our Theorem 8 for agents with ρ being entropic risk (cf. Example 1): *If each trader i has risk tolerance b_i and a belief parameterized by θ_i, and the initial market state is θ_0, then the equilibrium state of the market, to which the market converges, is given by $\theta^* = \frac{\theta_0 + \sum_i b_i \theta_i}{1 + \sum_i b_i}$.*

As the focus of this paper is on the convergence, the details for this result are given in Appendix C.

The main insight that drives the above analysis of the interaction between a risk-based trader and a market maker is that each trade minimizes a global objective for the market that is the *infimal convolution* [6] of the traders' and market maker's risks. In fact, this observation naturally generalizes to trades between three or more agents and the same convergence analysis applies. In other words, our analysis also holds when bilateral trade with a fixed market maker is replaced by multilateral trade among arbitrarily overlapping subsets of agents. Viewed as a graph with agents as nodes, the standard prediction market framework is represented by the star graph, where the central market market interacts with traders sequentially and individually. More generally we have what we call a *trading network*, in which the structure of trades can form arbitrary connected graphs or even hypergraphs. An obvious choice is the complete graph, which can model a *decentralized* market, and in fact we can even compare the convergence rate of our dynamics between the centralized and decentralized models; see Appendix D.2 and the discussion in § 5.

4

3 General Trading Dynamics

The previous section described the two agent case of what is more generally known as the *optimal risk allocation problem* [6] where two or more agents express their preferences for positions via risk measures. This is formalized by considering N agents with risk measures $\rho_i : \mathcal{R} \to \mathbb{R}$ for $i \in [N] := \{1, \ldots, N\}$ who are asked to split a position $r \in \mathcal{R}$ in to per-agent positions $r_i \in \mathcal{R}$ satisfying $\sum_i r_i = r$ so as to minimise the total risk $\sum_i \rho_i(r_i)$. They note that the value of the total risk is given by the *infimal convolution* $\wedge_i \rho_i$ of the individual agent risks — that is,

$$(\wedge_i \rho_i)(r) := \inf \left\{ \sum_i \rho_i(r_i) : \sum_i r_i = r , \ r_i \in \mathcal{R} \right\}. \tag{2}$$

A key property of the infimal convolution, which will underly much of our analysis, is that its convex conjugate is the sum of the conjugates of its constituent functions. See *e.g.* [23] for a proof.

$$(\wedge_i \rho_i)^* = \sum_{i \in [N]} \rho_i^* . \tag{3}$$

One can think of $\wedge_i \rho_i$ as the "market risk", which captures the risk of the entire market (*i.e.*, as if it were a single risk-based agent) as a function of the net position $\sum_i r_i$ of its constituents. By definition, eq. (2) says that the market is trying to reallocate the risk so as to minimize this net risk. This interpretation is confirmed by eq. (3) when we interpret the duals as penalty functions as above: the penalty of π is the sum of the penalties of the market participants.

As alluded to above, we allow our agents to interact round by round by conducting trades, which are simply the exchange of outcome-contingent securities. Since by assumption our position space $\mathcal{R}$ is closed under linear combinations, a trade between two agents is simply a position which is added to one agent and subtracted from another. Generalizing from this two agent interaction, a trade among a set of agents $S \subseteq [N]$ is just a collection of trade vectors, one for each agent, which sum to 0. Formally, let $S \subseteq [N]$ be a subset of agents. A *trade on S* is then a vector of positions $dr \in \mathcal{R}^N$ (*i.e.*, a matrix in $\mathbb{R}^{N \times k}$) such that $\sum_{i \in S} dr_i = 0 \in \mathcal{R}$ and $dr_i = 0$ for all $i \notin S$. This last condition specifies that agents not in S do not change their position.

A key quantity in our analysis is a measure of how much the total risk of a collection of traders drops due to trading. Given some subset of traders S, the *S-surplus* is a function $\Phi_S : \mathcal{R}^N \to \mathbb{R}$ defined by $\Phi_S(r) = \sum_{i \in S} \rho_i(r_i) - (\wedge_i \rho_i)(\sum_{i \in S} r_i)$ which measures the maximum achievable drop in risk (since $\wedge_i \rho_i$ is an infimum). In particular, $\Phi(r) := \Phi_{[N]}(r)$ is the *surplus function*. The trades that achieve this optimal drop in risk are called *efficient*: given current state $r \in \mathcal{R}^N$, a trade $dr \in \mathcal{R}^N$ on $S \subseteq [N]$ is *efficient* if $\Phi_S(r + dr) = 0$.

Our following key result shows that efficient trades have remarkable structure: once the state r and subset S is specified, there is a unique efficient trade, up to cash transfers. In other words, the surplus is removed from the position vectors and then redistributed as cash to the traders; the choice of trade is merely in how this redistribution takes place. The fact that the derivatives match has strong intuition from prediction markets: agents must agree on the price.[2] The proof is in Appendix A.1.

Theorem 2. *Let $r \in \mathcal{R}^N$ and $S \subseteq [N]$ be given.*

  i. *The surplus is always finite: $0 \leq \Phi_S(r) < \infty$.*

 ii. *The set of efficient trades on S is nonempty.*

iii. *Efficient trades are unique up to zero-sum cash transfers: Given efficient trades $dr^*, dr \in \mathcal{R}^N$ on S, we have $dr = dr^* + (z_1 r^{\$}, \ldots, z_N r^{\$})$ for some $z \in \mathbb{R}^N$ with $\sum_i z_i = 0$.*

 iv. *Traders agree on "prices": A trade dr on S is efficient if and only if for all $i, j \in S$, $\nabla \rho_i(r_i + dr_i) = \nabla \rho_j(r_j + dr_j)$.*

  v. *There is a unique "efficient price": If dr is an efficient trade on S, for all $i \in S$ we have $\nabla \rho_i(r_i + dr_i) = -\pi_S^*$, where $\pi_S^* = \underset{\pi \in \Pi}{\arg\min} \sum_{i \in S} \alpha_i(\pi) - \langle \pi, \sum_{i \in S} r_i \rangle$.*

---

[2]As intuition for the term "price", consider that the highest price-per-unit agent i would be willing to pay for an infinitesimal quantity of a position dr_i is $dr_i \cdot (-\nabla \rho_i(r_i))$, and likewise the lowest price-per-unit to sell. Thus, the entries of $-\nabla \rho_i(r_i)$ act as the "fair" prices for their corresponding basis positions/securities.

The above properties of efficient trades drive the remainder of our convergence analysis of network dynamics. It also allows us to write a simple closed form for the market price when traders share a common risk profile (Theorem 8). Details are in Appendix C. Beyond our current focus on rates, Theorem 2 has implications for a variety of other economic properties of trade networks. For example, in Appendix B we show that efficient trades correspond to fixed points for more general dynamics, market clearing equilibria, and equilibria of natural bargaining games among the traders.

Recall that in the prediction market framework of [13], each round has a single trader, say $i > 1$, interacting with the market maker who we will assume has index 1. In the notation just defined this corresponds to choosing $S = \{1, i\}$. We now wish to consider richer dynamics where groups of two or more agents trade efficiently each round. To this end will we call a collection $\mathcal{S} = \{S_j \subseteq [N]\}_{j=1}^m$ of groups of traders a *trading network* and assume there is some fixed distribution D over $\mathcal{S}$ with full support. A *trade dynamic* over $\mathcal{S}$ is a process that begins at $t = 0$ with some initial positions $r^0 \in \mathcal{R}^N$ for the N traders, and at each round t, draws a random group of traders $S^t \in \mathcal{S}$ according to D, selects some efficient trade dr^t on S, then updates the trader positions using $r^{t+1} = r^t + dr^t$.

For the purposes of proving the convergence of trade dynamics, a crucial property is whether all traders can directly or indirectly affect the others. To capture this we will say a trade network is *connected* if the hypergraph on $[N]$ with edges given by $\mathcal{S}$ is connected; *i.e.*, information can propagate throughout the entire network. Dynamics over classical prediction markets are always connected since any pair of groups from its network will always contain the market maker.

4 Convergence Analysis of Randomized Subspace Descent

Before briefly reviewing the literature on coordinate descent, let us see why this might be a useful way to think of our dynamics. Recall that we have a set $\mathcal{S}$ of subsets of agents, and that in each step, an efficient trade dr is chosen which only modifies the positions of agents in the sampled $S \in \mathcal{S}$. Thinking of $(r_1, \ldots, r_N)$ as a vector of dimension $N \cdot k$ vector (recall $\mathcal{R} = \mathbb{R}^k$), changing r^t to $r^{t+1} = r^t + dr$ thus only modifies $|S|$ blocks of k entries. Moreover, efficiency ensures that dr *minimizes* the sum of the risks of agents in S. Hence, ignoring for now the constraint that the sum of the positions must remain constant, the trade dynamic seems to be performing a kind of block coordinate descent of the surplus function Φ.

4.1 Randomized Subspace Descent

Several randomized coordinate descent methods have appeared in the literature recently, with increasing levels of sophistication. While earlier methods focused on updates which only modified disjoint blocks of coordinates [18, 22], more recent methods allow for more general configurations, such as overlapping blocks [17, 16, 20]. In fact, these last three methods are closest to what we study here; the authors consider an objective which decomposes as the sum of convex functions on each coordinate, and study coordinate updates which follow a graph structure, all under the constraint that coordinates sum to 0. Despite the similarity of these methods to our trade dynamics, we require even more general updates, as we allow coordinate i to correspond to arbitrary subsets $S_i \in \mathcal{S}$.

Instead, we establish a unification of these methods which we call *randomized subspace descent (RSD)*, listed in Algorithm 1. Rather than blocks of coordinates or specific linear constraints, RSD abstracts away these constructs by simply specifying "coordinate subspaces" in which the optimization is to be performed. Specifically, the algorithm takes a list of projection matrices $\{\Pi_i\}_{i=1}^n$ which define the subspaces, and at each step t selects a Π_i at random and tries to optimize the objective under the constraint that it may only move within the image space of Π_i; that is, if the current point is x^t, then $x^{t+1} - x^t \in \text{im}(\Pi_i)$.

Before stating our convergence results for Algorithm 1, we will need a notion of smoothness relative to our subspaces. Specifically, we say F is L_i-Π_i-*smooth* if for all i there are constants $L_i > 0$ such that for all $y \in \text{im}(\Pi_i)$,

$$F(x + y) \leq F(x) + \langle \nabla F(x), y \rangle + \tfrac{L_i}{2} \|y\|_2^2 . \tag{4}$$

Finally, let $F^{\min} := \min_{y \in \text{span}\{\text{im}(\Pi_i)\}_i} F(x^0 + y)$ be the global minimizer of F subject to the constraints from the Π_i. Then we have the following result for a constant $\mathcal{R}(x^0)$ which increases in:

ALGORITHM 1: Randomized Subspace Descent

Input: Smooth convex function $F : \mathbb{R}^n \to \mathbb{R}$, initial point $x^0 \in \mathbb{R}^n$, matrices $\{\Pi_i \in \mathbb{R}^{n \times n}\}_{i=1}^m$,
smoothness parameters $\{L_i\}_{i=1}^m$, distribution $p \in \Delta_m$
  for iteration t in $\{0, 1, 2, \cdots\}$ **do**
    sample i from p
    $x^{t+1} \leftarrow x^t - \frac{1}{L_i} \Pi_i \nabla F(x^t)$
end

---

(1) the distance from the point x^0 to furthest minimizer of F, (2) the Lipschitz constants of F w.r.t. the Π_i, and (3) the connectivity of the hypergraph induced by the projections.

Theorem 3. *Let F, $\{\Pi_i\}_i$, $\{L_i\}_i$, x^0, and p be given as in Algorithm 1, with the condition that F is L_i-Π_i-smooth for all i. Then $\mathbb{E}\left[F(x^t) - F^{\min}\right] \leq 2\mathcal{R}^2(x^0) / t$.*

The proof is in Appendix D. Additionally, when F is strongly convex, meaning it has a uniform local quadratic lower bound, RSD enjoys faster, linear convergence. Formally, this condition requires F to be μ-strongly convex for some constant $\mu > 0$, that is, for all $x, y \in \text{dom } F$ we require

$$F(y) \geq F(x) + \nabla F(x) \cdot (y - x) + \tfrac{\mu}{2}\|y - x\|^2 . \tag{5}$$

The statement and details of this stronger result is given in Appendix D.1.

Importantly for our setting these results only track the progress per iteration. Thus, they apply to more sophisticated update steps than a simple gradient step as long as they improve the objective by at least as much. For example, if in each step the algorithm computed the exact minimizer $x^{t+1} = \arg\min_{y \in \text{im}(\Pi_i)} F(x^t + y)$, both theorems would still hold.

4.2 Convergence Rates for Trade Dynamics

To apply Theorem 3 to the convergence of trading dynamics, we let $F = \Phi$ and $x = (r_1, \ldots, r_N) \in \mathcal{R}^N \cong \mathbb{R}^{Nk}$ be the joint position of all agents. For each subset $S \in \mathcal{S}$ of agents, we have a subspace of $\mathcal{R}^N$ consisting of all possible trades on S, namely $\{dr \in \mathcal{R}^N : dr_i = 0 \text{ for } i \neq S, \sum_{i \in S} dr_i = 0\}$, with corresponding projection matrix Π_S. For the special case of prediction markets with a centralized market maker, we have $N - 1$ subspaces $\mathcal{S} = \{\{1, i\} : i \in \{2, \ldots, N\}\}$ and $\Pi_{1,i}$ projects onto $\{dr \in \mathcal{R}^N : dr_i = -dr_1, dr_j = 0 \text{ for } j \neq 1, i\}$. The intuition of coordinate descent is clear now: the subset S of agents seek to minimize the total surplus within the subspace of trades on S, and thus the coordinate descent steps of Algorithm 1 will correspond to roughly efficient trades.
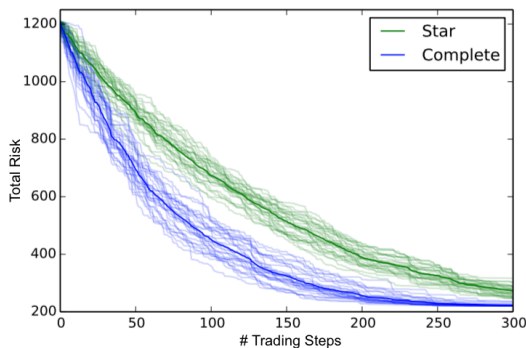
We now apply Theorem 3 to show that trade dynamics achieve surplus $\epsilon > 0$ in time $O(1/\epsilon)$. Note that we will have to assume the risk measure ρ_i of agent i is L_i-smooth for some $L_i > 0$. This is a very loose restriction, as our risk measures are all differentiable by the expressiveness condition.

Theorem 4. *Let ρ_i be an L_i-smooth risk measure for all i. Then for any connected trade dynamic, we have $\mathbb{E}\left[\Phi(r^t)\right] = O(1/t)$.*

Proof. Taking $L_S = \max_{i \in S} L_i$, one can check that F is L_S-Π_S-smooth for all $S \in \mathcal{S}$ by eq. (4). Since Algorithm 1 has no state aside from x^t, and the proof of Theorem 3 depends only the drop in F per step, any algorithm selecting the sets $S \in \mathcal{S}$ with the same distribution and satisfying $F(x^{t+1}) \leq F(x^t - \frac{1}{L_i}\Pi_i \nabla F(x^t))$ will yield the same convergence rate. As trade dynamics satisfy $F(x^{t+1}) = \min_{y \in \mathbb{R}^{Nk}} F(x^t - \Pi_i y)$, this property trivially holds, and so Theorem 3 applies. $\square$

If we assume slightly more, that our risk measures have local quadratic lower bounds, then we can obtain linear convergence. Note that this is also a relatively weak assumption, and holds whenever the risk measure has a Hessian with only one zero eigenvalue (for $r^\$$) at each point. This is satisfied, for example, by all the variants of entropic risk we discuss in the paper. The proof is in Appendix D.

Theorem 5. *Suppose for each i we have a continuous function $\mu_i : \mathcal{R} \to \mathbb{R}_+$ such that for all r, risk ρ_i is $\mu_i(r)$-strongly convex with respect to $r^{\$\perp}$ in a neighborhood of r; in other words, eq. (5) holds for $F = \rho_i$, $\mu = \mu_i(r)$, and all y in a neighborhood of r such that $(r - y) \cdot r^\$ = 0$. Then for all connected trade dynamics, $\mathbb{E}\left[\Phi(r^t)\right] = O(2^{-t})$.*

Table 1: Algebraic connectivities for common graphs.

Graph	$\|V(G)\|$	$\|E(G)\|$	$\lambda_2(G)$
K_n	n	$n(n-1)/2$	n
P_n	n	$n-1$	$2(1-\cos\frac{\pi}{n})$
C_n	n	n	$2(1-\cos\frac{2\pi}{n})$
$K_{\ell,k}$	$\ell+k$	ℓk	k
B_k	2^k	$k2^{k-1}$	2

Figure 1: Average (in bold) of 30 market simulations for the complete and star graphs. The empirical gap in iteration complexity is just under 2 (cf. Fig. 3).

Amazingly, the convergence rates in Theorem 4 and Theorem 5 hold for *all* connected trade dynamics. The constant hidden in the $O(\cdot)$ does depend on the structure of the network but can be explicitly determined in terms its *algebraic connectivity*. This is discussed further in Appendix D.2.

The intuition behind these convergence rates given here is that agents in whichever group S is chosen always trade to fully minimize their surplus. Because the proofs (in Appendix D) of these methods merely track the reduction in surplus per trading round, the bounds apply as long as the update is at least as good as a gradient step. In fact, we can say even more: if only an ϵ fraction of the surplus is taken at each round, the rates are still $O(1/(\epsilon t))$ and $O((1-\epsilon\mu)^t)$, respectively. This suggests that our convergence results are robust with respect to the model of rationality one employs; if agents have bounded rationality and can only compute positions which approximately minimize their risk, the rates remain intact (up to constant factors) as long as the inefficiency is bounded.

5   Conclusions & Future Work

Using the tools of convex analysis to analyse the behavior of markets allows us to make precise, quantitative statements about their global behavior. In this paper we have seen that, with appropriate assumptions on trader behaviour, we can determine the rate at which the market will converge to equilibrium prices, thereby closing some open questions raised in [2] and [13].

In addition, our newly proposed trading networks model allow us to consider a variety of prediction market structures. As discussed in §3, the usual prediction market setting is *centralized*, and corresponds to a star graph with the market maker at the center. A *decentralized* market where any trader can trade with any other corresponds to a complete graph over the traders. We can also model more exotic networks, such as two or more market maker-based prediction markets with a risk minimizing arbitrageur or small-world networks where agents only trade with a limited number of "neighbours".

Furthermore, because these arrangements are all instances of trade networks, we can immediately compare the convergence rates across various constraints on how traders may interact. For example, in Appendix D.2, we show that a market that trades through a centralized market maker incurs an quantifiable efficiency overhead: convergence takes twice as long (see Figure 1). More generally, we show that the rates scale as $\lambda_2(G)/|E(G)|$, allowing us to make similar comparisons between arbitrary networks; see Table 1. This raises an interesting question for future work: given some constraints such as a bound on how many traders a single agent can trade with, the total number of edges, etc, which network optimizes the convergence rate of the market? These new models and the analysis of their convergence may provide new principles for building and analyzing distributed systems of heterogeneous and self-interested learning agents.

References

[1] Jacob Abernethy, Yiling Chen, and Jennifer Wortman Vaughan. Efficient market making via convex optimization, and a connection to online learning. *ACM Transactions on Economics and Computation*, 1(2):12, 2013.

[2] Jacob Abernethy, Sindhu Kutty, Sébastien Lahaie, and Rahul Sami. Information aggregation in exponential family markets. In *Proceedings of the fifteenth ACM conference on Economics and computation*, pages 395–412. ACM, 2014.

[3] Jacob D Abernethy and Rafael M Frongillo. A collaborative mechanism for crowdsourcing prediction problems. In *Advances in Neural Information Processing Systems*, pages 2600–2608, 2011.

[4] Aharon Ben-Tal and Marc Teboulle. An old-new concept of convex risk measures: The optimized certainty equivalent. *Mathematical Finance*, 17(3):449–476, 2007.

[5] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[6] Christian Burgert and Ludger Rüschendorf. On the optimal risk allocation problem. *Statistics & decisions*, 24(1/2006):153–171, 2006.

[7] Yiling Chen and Jennifer Wortman Vaughan. A new understanding of prediction markets via no-regret learning. In *Proceedings of the 11th ACM conference on Electronic commerce*, pages 189–198. ACM, 2010.

[8] Nair Maria Maia de Abreu. Old and new results on algebraic connectivity of graphs. *Linear algebra and its applications*, 423(1):53–73, 2007.

[9] Hans Föllmer and Alexander Schied. *Stochastic Finance: An Introduction in Discrete Time*, volume 27 of *de Gruyter Studies in Mathematics*. Walter de Gruyter & Co., Berlin, 2nd edition, 2004.

[10] Rafael M Frongillo, Nicolás Della Penna, and Mark D Reid. Interpreting prediction markets: a stochastic approach. In *Proceedings of Neural Information Processing Systems*, 2012.

[11] P.D. Grünwald and A.P. Dawid. Game theory, maximum entropy, minimum discrepancy and robust Bayesian decision theory. *The Annals of Statistics*, 32(4):1367–1433, 2004.

[12] JB Hiriart-Urruty and C Lemaréchal. Grundlehren der mathematischen wissenschaften. *Convex Analysis and Minimization Algorithms II*, 306, 1993.

[13] Jinli Hu and Amos Storkey. Multi-period trading prediction markets with connections to machine learning. In *Proceedings of the 31st International Conference on Machine Learning (ICML)*, 2014.

[14] Jono Millin, Krzysztof Geras, and Amos J Storkey. Isoelastic agents and wealth updates in machine learning markets. In *Proceedings of the 29th International Conference on Machine Learning (ICML-12)*, pages 1815–1822, 2012.

[15] Bojan Mohar. The Laplacian spectrum of graphs. In *Graph Theory, Combinatorics, and Applications*, 1991.

[16] I Necoara, Y Nesterov, and F Glineur. A random coordinate descent method on large-scale optimization problems with linear constraints. *Technical Report*, 2014.

[17] Ion Necoara. Random coordinate descent algorithms for multi-agent convex optimization over networks. *Automatic Control, IEEE Transactions on*, 58(8):2001–2012, 2013.

[18] Yurii Nesterov. Efficiency of coordinate descent methods on huge-scale optimization problems. *SIAM Journal on Optimization*, 22(2):341–362, 2012.

[19] Mindika Premachandra and Mark Reid. Aggregating predictions via sequential mini-trading. In *Asian Conference on Machine Learning*, pages 373–387, 2013.

[20] Sashank Reddi, Ahmed Hefny, Carlton Downey, Avinava Dubey, and Suvrit Sra. Large-scale randomized-coordinate descent methods with non-separable linear constraints. *arXiv preprint arXiv:1409.2617*, 2014.

[21] Mark D Reid, Rafael M Frongillo, Robert C Williamson, and Nishant Mehta. Generalized mixability via entropic duality. In *Proc. of Conference on Learning Theory (COLT)*, 2015.

[22] Peter Richtárik and Martin Takáč. Iteration complexity of randomized block-coordinate descent methods for minimizing a composite function. *Mathematical Programming*, 144(1-2):1–38, 2014.

[23] R.T. Rockafellar. *Convex analysis*. Princeton University Press, 1997.

[24] Amos J Storkey. Machine learning markets. In *International Conference on Artificial Intelligence and Statistics*, pages 716–724, 2011.

A Proofs of Results

A.1 Proof of Theorem 2

Proof. We begin with (iv). By [12, eq. X1.3.4.5], which gives a very general result about infimal convolutions, we have that the condition $\Phi_S(r + dr) = 0$ implies the existence of some π such that $\nabla \rho_i(r_i + dr_i) = \pi$ for all $i \in S$. Conversely, we can appeal to [12, Prop X1.3.4.2], which again is for general infimal convolutions, to conclude $\Phi(r + dr) = 0$.

Now for $y \in \mathcal{R}$ consider $F(y) := \left(\wedge_{i \in S} \rho_i \right)(y)$, and let $x = \sum_{i \in S} r_i$. By eq. (3) and the definition of the conjugate, we have $\pi = -\nabla F(x) = \pi_S^*$ as defined in (v). Turning to (i), note that as $\text{relint}\Pi \subseteq \text{dom} \, \rho_i^*$ for all i by the expressiveness condition, we have $\text{relint}\Pi \subseteq \cap_{i \in S} \text{dom} \, \rho_i^* = \text{dom} \, F^*$. Now [12, Prop X1.3.4.1, eq. (XI.3.4.2)] gives us some $r' \in \mathcal{R}^N$ with $\sum_{i \in S} r'_i = x$, and $r'_i = r_i$ for $i \notin S$, such that $F(x) = \sum_{i \in S} \rho_i(r'_i)$, giving us $\Phi_S(r) = \sum_{i \in S} \rho_i(r_i) - F(x) = \sum_{i \in S} \rho_i(r_i) - \sum_{i \in S} \rho_i(r'_i) < \infty$. As $\Phi_S(r) \geq 0$ by definition, we have (i), and taking $dr = r' - r$, which is a trade as $\sum_i dr_i = \sum_{i \in S} r' - \sum_{i \in S} r_i + \sum_{i \notin S} (r'_i - r_i) = x - x + 0$, we also have (ii).

We can now settle (iii): by cash invariance, it is clear that $\nabla \rho_i(r_i + dr_i) = \nabla \rho_i(r_i + dr_i + z_i r^\$)$ for all $z_i \in \mathbb{R}$, and the strict risk aversion property says that these are the only such positions with the same derivative (otherwise convexity would imply ρ_i is flat in between, a contradiction). The requirement that $\sum_i dr_i = 0$ ensures $\sum_i z_i = 0$. $\qquad \square$

A.2 Proof of Theorem 8

Proof. Working with compressed positions, we have $\rho_i(r_i) = b\rho(\theta_i + r_i/b) - b\rho(\theta_i)$, where we have overloaded $\rho(r_i) = \rho(X[r_i])$. Taking $r_i^0 = 0$ for all i, by Theorem 7 and Theorem 2, the market clearing price is the unique price such that $\nabla \rho_i(dr_i) = \pi^*$ for all i, for some trade $dr \in \mathcal{R}^N$. By definition of ρ_i we have $\nabla \rho_i(dr_i) = \nabla \rho(\theta_i + dr_i/b_i)$. Now letting $\bar{b} = \sum_{j=1}^N b_j$, we see that taking $dr_i = b_i \sum_{j=1}^N b_j \theta_j / \bar{b} - b_i \theta_i$ gives us a valid trade with $\sum_i dr_i = \bar{b} \sum_{j=1}^N b_j \theta_j / \bar{b} - \sum_{i=1}^N b_i \theta_i = 0$, and by symmetry we have $\pi^* = \nabla \rho(\sum_{j=1}^N b_j \theta_j / \bar{b})$, meaning $\theta^* = \sum_{j=1}^N b_j \theta_j / \bar{b}$ as desired. $\qquad \square$

B Bargaining and Equilibria in Trade Networks

The key result of Theorem 2 concerning efficient trades provide a wealth of structure to help us understand market behaviour. The next result shows that efficient trades are fixed points of a dynamic. We say a state $r \in \mathcal{R}^N$ is a *fixed point* of a dynamic D if $r^t = r$ implies $r^s = r$ for all $s > t$.

Theorem 6. *Let $D(\mathcal{S}, p)$ be a connected trade dynamic. Then $r \in \mathcal{R}^N$ is a fixed point of D if and only if it is efficient.*

Proof. If $\Phi(r) = 0$, then $dr = 0$ is the only efficient monotone trade, so r is a fixed point of D. Conversely, if r is a fixed point of D, then as p has full support, we must have $\Phi_S(r) = 0$ for all $S \in \mathcal{S}$; otherwise with constant probability we would have $S = S^t$ for some t and thus some efficient trade $dr^t \neq 0$ and $r^{t+1} \neq r$. By Theorem 2 (iv), this means that for all $S \in \mathcal{S}$ and all $i, j \in S$ we have $\nabla \rho_i(r_i) = \nabla \rho_j(r_j)$. This gives us an equivalence classes of derivatives for each S, and by connectedness $\mathcal{S}$, the equivalence classes in fact must coincide. Thus, we have some π for which $\nabla \rho_j(r_j) = \pi$ for all $j \in [N]$, and again by Theorem 2 (iv) we conclude $\Phi(r) = 0$. $\qquad \square$

We observe that the combination of Theorems 2 and 6 means that for connected trade dynamics D, the fixed point of D corresponds to the efficient price π^* and hence the efficient trade for the overall market.

The result of Theorem 6 is somewhat surprising — not only is there a unique equilibrium (up to cash transfers) for all connected dynamics, but all connected dynamics have the *same* equilibrium! If one restricts to connected graphical networks, this means that the equilibrium does not depend on the network structure. The power of our framework is that the equilibrium analysis holds regardless of the way agents interact, as long as information is allowed to spread to all agents eventually. In fact, one could even consider an arbitrary process choosing subsets S^t of agents to trade at each time t; if

the set $\mathcal{S}$ of subsets which are visited infinitely often yields a connected hypergraph, then the proof Theorem 6 still applies.

We now explore two other types of equilibria in this setting. The first is a natural game-theoretic equilibrium where no group of agents can trade for mutual gain. The second, in the following subsection, looks at the classical market clearing condition in this setting.

Proposition 1. *Given connected $\mathcal{S}$, a point r is a game-theoretic equilibrium, meaning there is no set of agents $S \in \mathcal{S}$ which can mutually benefit by trading, if and only if r is efficient.*

Proof. Following the same logic as the proof of Theorem 6, we first argue that $\Phi_S(r) = 0$ for all S, and by connectedness this gives us some π such that $\nabla \rho_i(r) = -\pi$ for all i, from which we conclude $\Phi(r) = 0$. $\qquad\square$

B.1 Market clearing

Viewing the basis positions $\phi(\cdot)_i$ as goods in a marketplace, and the initial (compact) positions $r_i^0 \in \mathcal{R}$ as the endowments of the agents, it is natural to ask if there is a set of prices for the goods which clears the market. That is, if we set some price vector $\pi \in \Pi$ such that each agent buys some vector dr_i at those prices, we wish to find π such that $\sum_i dr_i = 0$.

Let us examine the optimization problem for each agent. Upon purchasing a bundle of goods $dr_i \in \mathcal{R}$ at prices π, for a total cost of $dr_i \cdot \pi$, the agent has net position $r_i^0 + dr_i - (dr_i \cdot \pi)r^\$$. In particular, it is now clear that we must have $r^\$ \cdot \pi = 1$; otherwise, if e.g. $r^\$ \cdot \pi < 1$, an agent purchasing $\lambda r^\$$ at cost would have final position $\lambda(1 - r^\$ \cdot \pi)r^\$$, which corresponds to an arbitrarily large risk-free payoff. Finally, as we simply factor the price of a bundle back into the position itself, the agent's purchasing decision is equivalent to choosing some dr_i such that $dr_i \cdot \pi = 0$.

As the utility of the agent is entirely captured by its risk measure ρ_i, we can clearly state the agent's optimization problem: given price vector π, agent i chooses a bundle of goods dr_i given by

$$dr_i \in \arg\min_{x:x\cdot\pi=0} \rho_i(r_i^0 + x) . \tag{6}$$

Via the method of Lagrange multipliers, a minimizing bundle dr_i in (6) must satisfy the constraint $\nabla \rho_i(r_i^0 + dr_i) = \lambda\pi$ for some $\lambda \in \mathbb{R}$. By cash invariance, we know that for all $\pi' \in -\nabla \rho_i(\cdot)$ we have $r^\$ \cdot \pi' = 1$, and as $r^\$ \cdot \pi = 1$ by the above argument, we must have $\lambda = 1$. We now have $\nabla \rho_i(r_i^0 + dr_i) = -\pi$ for all i. If we assume that π clears the market, then we additionally have $\sum_i dr_i = 0$, meaning dr is a valid trade, and thus Theorem 2 implies $\Phi(r^0 + dr) = 0$. Again by Theorem 2, we conclude that in fact $\pi = \pi^* = \arg\min_{\pi\in\Pi} \sum_i \alpha_i(\pi) - \langle \pi, \sum_i r_i^0 \rangle$.

Of course, it remains to show that the market clearing condition can be satisfied given the constraint $dr_i \cdot \pi^* = 0$. By Theorem 2 (ii), we know there exists an efficient trade dr', and by (iii) it is unique up to zero-sum cash transfers. Thus, taking $dr_i = dr_i' - (dr_i' \cdot \pi^*)r^\$$, which is zero-sum as $\sum_i dr_i' = 0$, we see that there is a unique optimal allocation $r = r^0 + dr$ to the agents, and which clears the market.

From the above discussion, we have the following result which refines the equilibrium concepts above; instead of a unique equilibrium up to zero-sum cash transfers, the market clearing allocation fixes a single position for each agent.

Theorem 7. *Let goods $\{\phi(\cdot)_i\}_{i=1}^k$, initial endowments $r^0 \in \mathcal{R}^N$, and (negative) utilities $\{\rho_i\}_{i=1}^N$ be given. Then the unique market clearing price π^* is given by Theorem 2 (v) and yields a unique and efficient allocation dr, which satisfies $dr_i \cdot \pi^* = 0$ for all i.*

Intuitively, the market clearing allocation gives us a benchmark for the proper way to "divide up the surplus" among the agents; given an efficient trade dr, perhaps the outcome of bargaining, each agent i should readjust by subtracting $dr \cdot \pi^*$ in cash.

C Optimal Allocation for Risk Compatible Traders

The price matching property of efficient trades (Theorem 2) allows us to find closed form solutions for the optimal allocation of risk in a market when the traders' risks are suitably "compatible".

These risk measures are derived from a *base* risk ρ which allows us to capture a notion of beliefs and varying levels of risk aversion, as follows. Let

$$\rho_{b,s}(r) := b\rho(s + r/b) - b\rho(s), \tag{7}$$

where $b > 0$ is the *risk tolerance* and $s \in \mathcal{R}$ is some reference position. We will call the risk measures $\rho_{b,s}$ *compatible* with ρ and can view them as a combination of a translation $\rho(r) = \rho(s + r) - \rho(s)$, which simulates the additional risk imposed by r having already taken on position s, and a *perspective transform* $\rho(r) = b\rho(r/b)$, where a higher b corresponds to a less risk-averse agent.[3]

In general, for an agent with risk measure $\rho : \mathcal{R} \to \mathbb{R}$, we can think of $-\nabla\rho(0)$ as the agent's belief. This can be justified in two ways: (1) when ρ is the (negative) certainty equivalent for exponential utility, for an agent with belief p_s, then $-\nabla\rho(0) = p_s$; (2) the penalty function α, which specifies how "surprising" each distribution $p \in \Delta_\Omega$ is, is minimized at $p = -\nabla\rho(0)$. For this reason, and the fact that $\nabla\rho_{b,s}(0) = \nabla\rho(s) = -p_s$, we may think of $\rho_{b,s}$ as the risk of an agent with initial belief p_s and risk tolerance b.

Consider agents with inital beliefs p_{s_i} and risk tolerances $b_i > 0$, meaning $\rho_i = \rho_{b_i,s_i}$. What is the final market price in terms of these parameters? Motivated by the prediction market setting, we would in particular like to see some sort of sensible *aggregation* of these beliefs. As we now show, the market clearing price corresponds to a risk-tolerance-weighted average of the agents' parameters.

Theorem 8. *Let ρ be a risk measure, and let each agent i have compatible risk measures ρ_{b_i,s_i} initial belief p_{s_i} and risk tolerance parameter $b_i > 0$. Then the market clearing price is given by p_{s^*}, where*

$$s^* = \frac{\sum_i b_i s_i}{\sum_i b_i} . \tag{8}$$

This result generalizes those in §5 of [2], where traders are assumed to maximize an expected utility of the form $U_b(w) = -b\exp(-w/b)$ under beliefs drawn from an exponential family with sufficient statistic given by the securities ϕ. The above result shows that exactly the same weighted distribution of positions at equilibrium occurs for *any* family of risk-based agents, not just those derived from exponential utility via certainty equivalents [4]. In addition, this generalization shows that the agents need not have exponential family beliefs: their positions θ_i act as general natural parameters, and $1/b_i$ acts as a general measure of risk aversion. Finally, applying Theorem 4 or 5 establishes convergence rates for this setting, which addresses the future work in [2].

D   Random Subspace Descent

For the analysis, we introduce a seminorm $\|\cdot\|_A$ which will measure the progress per iteration, and its dual $\|\cdot\|_A^*$.

$$\|x\|_A := \left(\sum_{i=1}^m \frac{p_i}{L_i} \|\Pi_i x\|_2^2 \right)^{1/2} . \tag{9}$$

$$\|y\|_A^* := \begin{cases} \langle A^+ y, y \rangle^{1/2} & \text{if } y \in \text{im}(A) \\ \infty & \text{otherwise.} \end{cases} \tag{10}$$

Note that $\|\cdot\|_A$ is a Euclidean seminorm $\|x\|_A = \langle Ax, x \rangle$ with $A = \sum_i \frac{p_i}{L_i} \Pi_i$. One can check that $\|\cdot\|_A^*$ is indeed the dual norm of $\|\cdot\|_A$, in the sense that $\left(\frac{1}{2}\|\cdot\|_A^2\right)^* = \frac{1}{2}\|\cdot\|_A^{*2}$, where the first $*$ denotes the convex conjugate.

Define $X(A) = \{x^0 + y : y \in \text{span}\{\text{im}(\Pi_i)\}_i\}$ to be the optimization domain, so that $F^{\min} = \min_{x \in X(A)} F(x)$, and let $F^{\text{arg}} := \arg\min_{x \in X(A)} F(x)$ denote the minimizers of F. Then we define the constant $\mathcal{R}(x^0)$ by

$$\mathcal{R}(x^0) := \max_{x \in X(A):F(x) \leq F(x^0)} \max_{x^* \in F^{\text{arg}}} \|x - x^*\|_A^* , \tag{11}$$

---

[3]Note that agents are never risk-seeking; only in the limit as $b \to \infty$ do the traders become risk-neutral.

the maximum distance, according to $\|\cdot\|_A^*$, between any minimizer of F and any feasible point at least as good as x^0.

We now have the foundation to prove Theorem 3.

Proof of Theorem 3. To begin, suppose subspace i is chosen at step t, and consider the update $x^{t+1} = x^t - y$ for $y \in \mathrm{im}(\Pi_i)$. The drop in the objective can be bounded using eq. (4),

$$F(x^t) - F(x^t - y) \geq \left\langle \nabla F(x^t), y \right\rangle - \frac{L_i}{2}\|y\|_2^2 \,. \tag{12}$$

By properties of orthogonal projections, we have

$$\arg\max_{y\in\mathrm{im}(\Pi_i)} \left\langle \nabla F(x^t), y \right\rangle - \frac{L_i}{2}\|y\|_2^2 \;=\; \arg\min_{y\in\mathrm{im}(\Pi_i)} \left\| y - \tfrac{1}{L_i}\nabla F(x^t) \right\|_2 \;=\; \tfrac{1}{L_i}\Pi_i\nabla F(x^t) \,,$$

and choice of y gives our update in Algorithm 1. Substituting this y into eq. (12) gives

$$F(x^t) - F(x^{t+1}) \geq \left\langle \nabla F(x^t), \tfrac{1}{L_i}\Pi_i\nabla F(x^t) \right\rangle - \frac{L_i}{2}\|\tfrac{1}{L_i}\Pi_i\nabla F(x^t)\|_2^2$$

$$= \frac{1}{2L_i}\|\Pi_i\nabla F(x^t)\|_2^2 \,.$$

Now looking at the expected drop in the objective, we have

$$F(x^t) - \mathbb{E}\left[F(x^{t+1})|x^t\right] \geq \sum_{i=1}^m p_i \frac{1}{2L_i}\|\Pi_i\nabla F(x^t)\|_2^2 = \frac{1}{2}\|\nabla F(x^t)\|_A^2 \,. \tag{13}$$

To relate our per-round progress to the gap remaining, we observe that

$$F(x^t) - F^{\min} \leq \max_{x^*\in\arg\min_x F(x)} \left\langle \nabla F(x^t), x^* - x^t \right\rangle$$

$$\leq \max_{x^*\in\arg\min_x F(x)} \|\nabla F(x^t)\|_A \, \|x^* - x^t\|_A^*$$

$$\leq \|\nabla F(x^t)\|_A \max_{x^*\in\arg\min F} \max_{x:F(x)\leq F(x^0)} \|x^* - x\|_A^*$$

$$= \|\nabla F(x^t)\|_A \, \mathcal{R}(x^0) \,,$$

where we used convexity of F, the definition of the dual norm, the fact that $F(x^t)$ is non-increasing in t, and finally the definition of $\mathcal{R}$. We now have $F(x^t) - \mathbb{E}\left[F(x^{t+1})|x^t\right] \geq (F(x^t) - F^{\min})^2/(2\mathcal{R}^2(x^0))$. The remainder of the proof follows an argument of [16] by analyzing $\Delta_t = \mathbb{E}\left[F(x^t) - F^{\min}\right]$. From the last inequality we have $\Delta_{t+1} \leq \Delta_t - \Delta_t^2/2\mathcal{R}^2(x^0)$, and since $0 \leq \Delta_{t+1} \leq \Delta_t$, dividing by $\Delta_t\Delta_{t+1}$ gives $\Delta_t^{-1} \leq \Delta_{t+1}^{-1} - (2\mathcal{R}^2(x^0))^{-1}$. Summing these inequalities gives the result. $\square$

D.1 Faster Rates for RSD

Theorem 9. *Let F, $\{\Pi_i\}_i$, $\{L_i\}_i$, x^0, and p be given as in Algorithm 1, with the condition that F is L_i-Π_i-smooth for all i, and additionally that F is μ-strongly convex with respect to $\|\cdot\|_A^*$. Then we have $\mathbb{E}\left[F(x^t) - F^{\min}\right] \leq (1-\mu)^t (F(x^0) - F^{\min})$.*

Proof of Theorem 9. Our proof is essentially that of Nesterov [18, Thm 2] and Richtárik and Takáč [22, Thm 12]. By definition of μ-strongly convex, we have for all $y \in \mathbb{R}^n$,

$$F(y) - F(x^t) \geq \left\langle \nabla F(x^t), y - x^t \right\rangle + \frac{\mu}{2}\|y - x^t\|_A^{*\,2} \,.$$

Independently minimizing each side of this inequality over y, we obtain from [22, Lemma 10],

$$F^{\min} - F(x^t) \geq -\tfrac{1}{2\mu}\|\nabla F(x^t)\|_A^2.$$

Now combining with eq. (13), we have

$$F(x^t) - \mathbb{E}\left[F(x^{t+1})|x^t\right] \geq \frac{1}{2}\|\nabla F(x^t)\|_A^2 \geq \mu(F(x^t) - F^{\min}) \,.$$

Taking expectations and rearranging, we have $\mathbb{E}\left[F(x^{t+1}) - F^{\min}\right] \leq (1-\mu)\mathbb{E}\left[F(x^t) - F^{\min}\right]$, from which the result follows by induction. $\square$

D.2 RSD for Graphs and Hypergraphs

In this section we consider a special case of Algorithm 1, where we have a linear constraint $\sum_i x_i = c$ on the coordinates, and the subspaces correspond to graphs (overlapping pairs of coordinates), or hypergraphs (overlapping subsets of coordinates).[4] In the graphical case, we will leverage existing results in spectral graph theory to analyze new graphs currently not considered in the literature. Note that we focus here on uniform probabilities to highlight the connections to spectral graph theory; for an analysis of the optimal probabilities, see Necoara et al. [16].

Let us first consider an optimization problem on the complete graph, which picks an edge (i, j) uniformly at random and optimizes in coordinates i and j under the constraint that $x_i^{t+1} + x_j^{t+1} = x_i^t + x_j^t$. One can check that this corresponds to the projection matrix $\Pi_{(i,j)} = \frac{1}{2}(e_i - e_j)(e_i - e_j)^\top$, where e_i is the ith standard unit vector. Assuming a global smoothness constant L, one can calculate

$$A = \tfrac{2}{Ln(n-1)} \sum_{(i,j)} \Pi_{(i,j)} = \tfrac{1}{L(n-1)} \left(I - \tfrac{1}{n} \mathbb{1}\mathbb{1}^\top \right) , \quad A^+ = L(n-1)(I - \tfrac{1}{n} \mathbb{1}\mathbb{1}^\top) ,$$

where $\mathbb{1}$ is the all-ones vector. Now as $\mathrm{im}(A) = \ker(\mathbb{1})$, this gives

$$\|x\|_A^{*\,2} = L(n-1)\|x\|_2^2 . \tag{14}$$

Similarly, the complete rank-k hypergraph gives $\|x\|_A^{*\,2} = L\frac{n-1}{k-1}\|x\|_2^2$. (Compare to eq. (3.10) and the top of p.21 of [16].) Letting $\mathcal{C}_0 = 4L \max_{x \in X(A): F(x) \leq F(x^0)} \max_{x^* \in F^{\mathrm{arg}}} \|x - x^*\|_2^2$, which is independent of the (hyper)graph as long as it is connected, we thus have a convergence rate of $\frac{n-1}{2}\mathcal{C}_0 \frac{1}{t}$ for the complete graph, and more generally $\frac{n-1}{2(k-1)}\mathcal{C}_0 \frac{1}{t}$ for the complete k-graph. Henceforth, we will consider the coefficient in front of $\mathcal{C}_0$ to be the convergence rate.

The above matrix A is a scaled version of what is known as the *graph Laplacian*; given graph G with adjacency matrix $A(G)$ and degree matrix $D(G)$ with the degrees of each vertex on the diagonal, the Laplacian is the matrix

$$\mathcal{L} = \mathcal{L}(G) := D(G) - A(G) . \tag{15}$$

One can check that indeed, $\mathcal{L} = 2\sum_{(i,j) \in E(G)} \Pi_{(i,j)}$, meaning $A = \frac{p}{2L}\mathcal{L}$, where $p = 1/|E(G)|$ is the uniform probability on edges.

The graph Laplacian is a well-studied object in spectral graph theory and other domains, and we can use existing results to establish bounds for other graphs of interest. To draw this connection, we note two facts: (1) for symmetric matrices B, the norm $\langle Bx, x \rangle^{1/2}$ can be bounded by the maximum eigenvalue of B, and (2) the maximum eigenvalue of B^+ is equal to the inverse of the smallest nonzero eigenvalue of B, provided again that B is symmetric.[5] Putting these together, we can therefore bound $\| \cdot \|_A^*$ using the smallest nonzero eigenvalue of A, and hence of $\mathcal{L}$. It is easy to see that the smallest eigenvalue is $\lambda_1(G) = 0$ with eigenvector $\mathbb{1}$, and as G is connected, we will have $\lambda_2(G) > 0$. Thus, the smallest nonzero eigenvalue of A is simply $\frac{p}{2L}\lambda_2(G)$, so we have the following for *any* connected graph G:

$$\|x\|_A^{*\,2} \leq 2L\frac{|E(G)|}{\lambda_2(G)}\|x\|_2^2 . \tag{16}$$

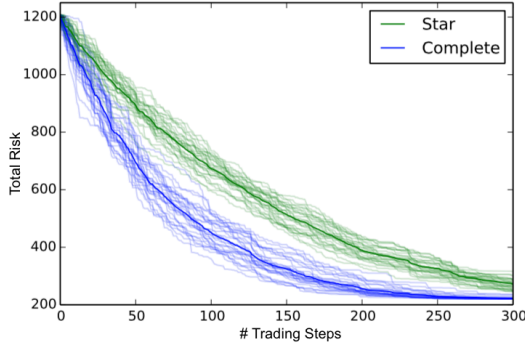Of course, by the above definition of $\mathcal{C}_0$ and Theorem 3, this yields the result

$$\mathbb{E}\left[F(x^t) - F^{\min} \right] \leq \frac{|E(G)|}{\lambda_2(G)}\mathcal{C}_0 \frac{1}{t} , \tag{17}$$

showing us how tightly related this eigenvalue is to rate of convergence of Algorithm 1.

The second-smallest eigenvalue $\lambda_2(G)$ is called the *algebraic connectivity* of G, and is itself thoroughly studied in spectral and algebraic graph theory. For example, it is known (and easy to check) that $\lambda_2(K_n) = n$, where K_n denotes the complete graph; this together with $|E(K_n)| = n(n-1)/2$ immediately gives eq. (14). In [8], algebraic connectivities are also given for the path on n vertices P_n, the cycle C_n, the bipartite complete graph $K_{\ell,k}$ for $k < \ell$, and the k-dimensional cube B_k. We collect these eigenvalues together yields Table 2.

---

[4]Everything in this section also holds for a graphical or hypergraphical structure on blocks of coordinates; just add Kronecker products with the appropriate identity matrix.

[5]These facts follow from the operator norm and singular-value decomposition for the pseudoinverse, respectively, together with the fact that singular values are eigenvalues for symmetric matrices.

Table 2: Algebraic connectivities for common graphs.

Graph	$\|V(G)\|$	$\|E(G)\|$	$\lambda_2(G)$
K_n	n	$n(n-1)/2$	n
P_n	n	$n-1$	$2(1-\cos\frac{\pi}{n})$
C_n	n	n	$2(1-\cos\frac{2\pi}{n})$
$K_{\ell,k}$	$\ell+k$	ℓk	k
B_k	2^k	$k2^{k-1}$	2

Figure 2: Average (in bold) of 30 runs of a separable objective for the complete and star graphs. The empirical gap in iteration complexity is just under 2 (cf. Fig. 3).

Substituting the values in Table 2 into eq. (16), we can directly compare the theoretical convergence rates for different graphs. For example, the star graph $K_{n-1,1}$ has rate $(n-1)(1)/(1) = (n-1)$, which is only a factor of 2 away from the complete graph.[6] The path and cycle fare much worse, yielding roughly $n/2(n^{-2}/2) = n^3$ as n becomes large (applying the Taylor expansion and ignoring π terms). Finally, an interesting result due to Mohar [15] says that for any connected graph on n vertices, we have $\lambda_2(G) \geq 4/(n\mathrm{diam}(G))$ where $\mathrm{diam}(G)$ is the diameter of G. Hence for any connected graph,

$$\mathbb{E}\left[F(x^t) - F^{\min}\right] \leq \frac{n\,|E(G)|\,\mathrm{diam}(G)}{4}\mathcal{C}_0\frac{1}{t}\,, \tag{18}$$

which is useful for sparse graphs of small diameter. See Appendix D.3 for more on hypergraphs.

D.3 Hypergraphs

Here we briefly show how to analyze general hypergraphs. Representing a hypergraph as a collection $\mathcal{S}$ of hyperedges $S \subseteq [n]$, we may define the degree matrix $D(\mathcal{S})$ to be the diagonal matrix with $D(\mathcal{S})_{ii} = \#\{S \in \mathcal{S} : i \in S\}$, and the "adjacency" matrix to be $A(\mathcal{S})_{ij} = \sum_{S \in \mathcal{S}: i,j \in S} 1/|S|$. Then for uniform probabilities we have $A = \frac{p}{L}(D(\mathcal{S}) - A(\mathcal{S}))$. This follows from observing that for subset S, we have $\Pi_S = I_S - \frac{1}{|S|}\mathbb{1}_S\mathbb{1}_S^\top$, and counting as we sum. Taking the complete k-graph yields $D(\mathcal{S}) = \binom{n-1}{k-1}I$ and $A(\mathcal{S})_{ij} = \frac{1}{k}\binom{n-2}{k-2} = \frac{k-1}{k(n-1)}\binom{n-1}{k-1}$ for $i \neq j$ and $A(\mathcal{S})_{ii} = \frac{1}{k}\binom{n-1}{k-1}$; putting these together gives $A = \frac{1}{L\binom{n}{k}}\frac{n}{k}\frac{k-1}{n-1}\binom{n-1}{k-1}(I - \frac{1}{n}\mathbb{1}\mathbb{1}^\top)) = \frac{n-1}{L(k-1)}(I - \frac{1}{n}\mathbb{1}\mathbb{1}^\top)$. Similar computations may be done for other hypergraphs of interest.

---

[6]While of course these are merely upper bounds on the true rates, they match Figures 2 and 3 quite well.
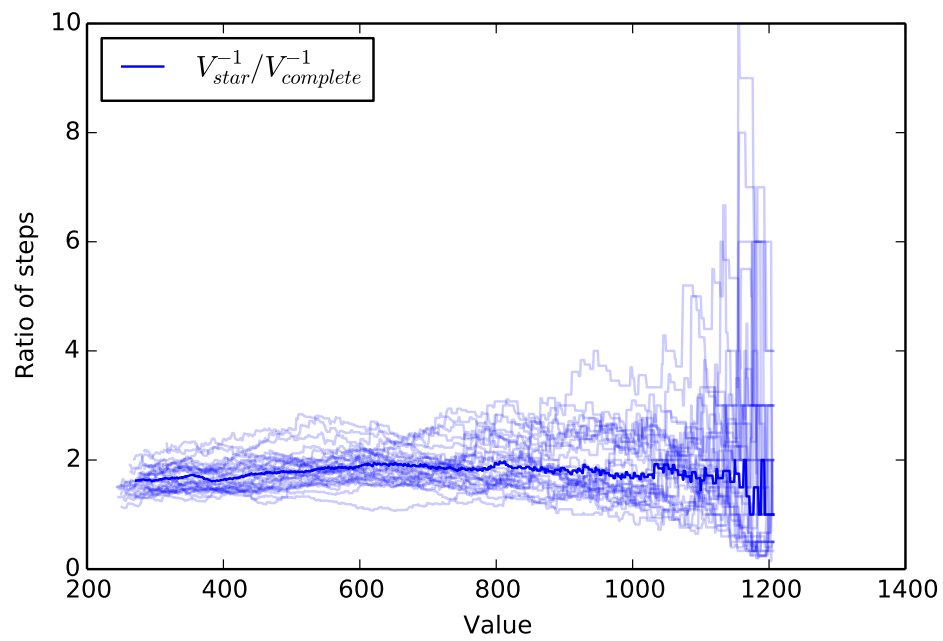
Figure 3: Thirty runs of a separable objective under the complete and star graphs. The ratio between star and complete of the number of iterations needed to achieve a given objective value is plotted, with the average in bold.