

EDIT: Thanks to great feedback by Glen Weyl, I've been able to significantly revise and extend this concept using the "bid language" framework. Please see [my newer post](#), although this one still serves as good background.

Quadratic Funding: Implementing the revelation principle using SGD

Abstract

: Quadratic funding is a mechanism which makes societally productive investments stable attractors ([visualization of dynamics](#)) while also providing a positive but unstable incentive for unproductive group investments (visualization of [dynamics](#), [payoff](#)). Although it has attractive equilibrium properties, playing the equilibrium is nontrivial. We argue that in this case, direct approximate revelation, provided the graphical interface is good, is simpler for users and hence further increases the gap between productive coordination difficulty and unproductive coordination difficulty. We describe an approximate elicitation procedure which follows the work of [Critch 2022](#), as well as a strategy executor which uses stochastic gradient descent (SGD), though we may use any global optimization algorithm such as [Adam](#), [Adagrad](#), etc. We then discuss the computational viability and how one may make the strategy executor trustless and/or private.

Informational challenges in standard QF strategy

[Buterin, Weyl, and Hitzig \(2018\)](#) note that "dynamic implementation is very likely to be desirable [...] as the optimal contribution will still depend on others' contributions". In other words, I do not know how much to contribute unless I know how much others contribute, but others do not know much to contribute unless they know much I contribute. Realistically, I will not be logging in very often to check and adjust my allocations.

Let us illustrate what may happen if naively show a static projection of match (which may misrepresent the final match). Then every project will start at a static match of \$0. The first project to receive a contribution will jump to a static match of 100% of the budget. This will fade over time, but the initial effect is so large that it may well persist, i.e. participants who are "baited in" will realistically not check back in and pull out.

The intuitive conclusion is "first movers" have disproportionate influence in the outcome of a public good (which may be exacerbated by first-mover advantages related to UI sort, etc). There is [relevant data from Bitcoin](#) which preliminarily suggests that there may indeed be a first mover advantage (determining causality would require further analysis).

Could we come up with a "smart" projection? Such an endeavor would likely face full-information constraints, introduce a centralized source of significant bias, or both.

Revelation Principle

The [revelation principle](#) states that for any mechanism m_1

there is a "direct mechanism" m_2

which directly asks agents for their preferences and executes the strategy for them. Whether m_2

is a more useful mechanism than m_1

depends on the description length of preferences ([combinatorial auctions are an example](#) of a setting where we generally do not want the direct mechanism). In the case of QF, the full preference description is a continuous function $V_i^p(F_p)$

; however, we may approximately

specify this function with just a few 1- or 2- dimensional datapoints. Note that we do not consider joint utility as that would blow up the description length; hence this works best for relatively independent goods or single-good games. ([see follow-up post](#))

Elicitation

The user interface is an endless opportunity for refinement, but a simple one might be a click/drag-based interface where one may draw and/or type a few values of V_i^p

. Another might be a "bulk prices" UI where the user bids against our "bulk discounts" until they are satisfied. We also explain to the user that the amount they spend on a project will never exceed the personal valuation that they are giving us. We would then use [any common interpolation method](#) and submit the resulting $V_{[i, \text{approx}]}^p$

to the strategy executor. Note that if the approximation of total marginal utility for a given project is off by a factor of k

, then the project may be underfunded or overfunded in the sense that marginal social benefit (MSB) would vary from other

projects by k

. We expect that as more people report for a given project, random mis-approximations will average out.

Another way to conceptualize this is that standard QF re-queries the user's bid every time they log in, since the match amount evolves. This UI queries a minimal set of bids upfront.

This scheme is similar to that presented in [Critch 2022](#), which also makes the case that a discussion component is more important than elicitation itself. Note that the setting of Critch 2022 is generally a highly dedicated subset of individuals which attempts to represent the utility of the whole world, rather than a wider group of less dedicated selfish agents (note that any effective public goods funding scheme likely has some properties of the former, since selflessness is required for projects benefitting non-constituents such as far-future or even not-so-far future citizens, and participation/dedication also follow an [80/20](#)). Critch's point makes sense, although further discussion is outside the scope of this main post.

Optimization

The user would authorize some payment to a bot, which would then run gradient descent, which simulates what a users might do if they were checking and adjusting allocations frequently.

Buterin, Weyl, and Hitzig (2018) also note that nonconcave utility is natural, but makes the system's attractors non-global (an example being a "cold start" utility curve - [visualization](#)). Hence, we would realize gains from using a stochastic

variant of gradient descent / global optimizer. In this sense, the SGD also acts as [a strategy modification inducing correlated equilibrium](#).

The usage of blockchain-SGD-coordinators in mechanisms in general could be a broader cryptoeconomic research area. For example, the agent's choice of loss function (e.g. [Rawlsian vs utilitarian](#)) affects [which equilibrium it chooses](#) (the [Folk Theorems](#) suggest that equilibrium selection will be important in many settings), which may give the designer [control over](#) efficiency vs equity, risk vs payoff, etc. A natural question would be whether a metagame emerges of people defecting between different coordinators of the same game.

The question of equilibrium selection relates to the curse of dimensionality - the number of attractors and the difficulty of discovering them may blow up in terms of the number of people. The fact that we are now optimizing deep learning models with over hundreds of billions parameters is comforting for us, although the success of optimization at that scale probably involves various cutting-edge techniques and manual tuning. We don't expect optimizing over thousands to millions of contribution amounts will be hugely problematic, especially if the space does turn out to be decently concave, and especially given that the nature of the problem is not a black box and we can make educated guesses about starting points. However, once we desire proofs, the computation suddenly gets a lot more expensive (given such constraints, consensus-level computation in the spirit of unvirtualized [Cosmos](#) blockchains, possibly with fraud proofs if they scale well enough, may be less ambitious starting ground). We stress that any optimization challenges faced by the SGD algorithm would be even more severe if mentally executed by participants (e.g. we argued that even in the simplest cases, participants are unlikely to discover equilibrium).

There may also be optimizations available from discrete optimization, if we use linear interpolation of the reported valuations. We need not descend along the gradient, but instead directly between points.

Interaction Complexity

The nature of public goods is such that every good i is relevant to every participant j . Hence, the transaction cost of the system is $X * Y * K$ parameters where X = the population, Y = the number of public good options, and K = the cost of transacting. We've argued that K is lower for entering a utility curve than it is for playing the direct contribution game, and is essentially minimal if the UI is good. The bottleneck is now $X * Y$. Downsampling X is likely to be unviable due to incentives, and improvements over random downsampling for Y are likely to come from peer-review, search, and personalized recommendation.

Consent, Trustlessness, and Privacy

Because we are authorizing bots to spend money on our behalf, we must enforce that the bot is doing the right thing. This is a use-case for consensus and/or proofs (though see earlier note on computation costs).

We also may desire that the inputs and individual allocations are private. The SGD approach is actually more

amenable to differential privacy (this is a use-case for [homomorphic threshold encryption](#)), since the computation is by nature maximally batched, i.e. the total final allocation doesn't have to be revealed until the very end. With direct contribution, we must update the allocation estimate periodically (technically, we could not, but then the game becomes truly one-shot and hence ~impossible to play), which exposes a tradeoff between update frequency and privacy.

Given this privacy capability, we may want to revisit [MACI anonymization](#).

Psychology

One could argue that the feeling of seeing a huge and/or inflated projected match is key to the enjoyment of participating in QF, that that is worth any potential distortions in the optimality of the mechanism, and/or that the optimality of the mechanism is an unrealistic goal in practice (e.g. “utility” is not a clear concept). If a large portion of the mechanism’s appeal is in fact its psychological effects rather than its game-theoretic effects, then we may still report projected matches (of course, this will reintroduce the update frequency tradeoff mentioned earlier).

Hybrid System

It is also possible to expose both the traditional QF experience and the SGD experience. In this case, it would be helpful for the SGD agent to communicate information to the traditional participants, which once again brings us to the update frequency tradeoff.