



Tax-Aware Portfolio Construction via Convex Optimization

Nicholas Moehle¹ · Mykel J. Kochenderfer² · Stephen Boyd³ · Andrew Ang⁴

Received: 29 September 2020 / Accepted: 27 January 2021 / Published online: 25 February 2021
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC part of Springer Nature 2021

Abstract

We describe an optimization-based tax-aware portfolio construction method that adds tax liability to standard Markowitz-based portfolio construction. Our method produces a trade list that specifies the number of shares to buy of each asset and the number of shares to sell from each tax lot held. To avoid wash sales (in which some realized capital losses are disallowed), we assume that we trade monthly and cannot simultaneously buy and sell the same asset. The tax-aware portfolio construction problem is not convex, but it becomes convex when we specify, for each asset, whether we buy or sell it. It can be solved using standard mixed-integer convex optimization methods at the cost of very long solve times for some problem instances. We present a custom convex relaxation of the problem that borrows curvature from the risk model. This relaxation can provide a good approximation of the true tax liability, while greatly enhancing computational tractability. This method requires the solution of only two convex optimization problems: the first determines whether we buy or sell each asset, and the second generates the final trade list. In our numerical experiments, our method almost always solves the nonconvex problem to optimality, and when it does not, it produces a trade list very close to optimal. Backtests show that the performance of our

Communicated by Klaus Reiner Schenk-Hoppe.

✉ Nicholas Moehle
nicholasmoehle@gmail.com

Mykel J. Kochenderfer
mykel@stanford.edu

Stephen Boyd
boyd@stanford.edu

Andrew Ang
andrew.ang@blackrock.com

¹ BlackRock AI Labs, BlackRock, Inc., Palo Alto, CA, USA

² Department of Aeronautics & Astronautics, Stanford University, Stanford, CA, USA

³ Department of Electrical Engineering, Stanford University, Stanford, CA, USA

⁴ Factor-Based Strategies Group, BlackRock, Inc., New York, NY, USA

method is indistinguishable from that obtained using a globally optimal solution, but with significantly reduced computational effort.

Keywords Portfolio optimization · Convex optimization · Convex relaxations · Shapley-Folkman lemma · Tax-aware portfolio management

1 Introduction

We formulate a tax-aware portfolio construction problem that explicitly accounts for tax liabilities from long- and short-term capital gains, while maintaining the standard objectives of return, risk, and transaction costs. While this tax liability term is not convex, we develop a convex relaxation of the problem that allows trade lists to be computed extremely quickly.

In addition to choosing how many shares to buy or sell of each asset, our method divides each sale across tax lots. For a fixed sell amount, we find it is optimal to sell shares in the order that minimizes immediate tax liability, which we call the least-tax-first-out (LTFO) method, coinciding with the well-known highest-basis-first-out (HIFO) method if the same tax rate applies to all lots. [2,17] show that the accounting method for lot ordering significantly affects the losses that can be harvested. Even though it is not optimal, [1] show that the HIFO method generates substantial benefit to an investor's total wealth.

Our convex relaxation combines all terms that are separable across a single asset, and then replaces the resulting (nonconvex) function with its convex envelope. This approach has the effect of taking (or 'borrowing') curvature from the transaction cost and specific risk terms and introducing it into the tax liability term. This effect can also be interpreted as an application of the Shapley–Folkman Lemma. The effect is that the tax-aware portfolio construction problem can be well approximated by a convex optimization problem, and the fidelity of this approximation is shown empirically.

To demonstrate our method, we apply it to a tax-loss harvesting strategy. First, we show that in a realistic backtest scenario, the strategy tightly tracks the benchmark (the S&P 500) while harvesting capital losses. We compare the performance of our method with that of a standard mixed-integer quadratic programming formulation and show that our method delivers near-identical trade lists, despite being several hundred times faster.

1.1 Related Work and Background

1.1.1 Markowitz Portfolio Construction

The formulation of portfolio construction as an optimization problem by [28] involves a trade-off of expected return and risk. This optimization problem, with a quadratic objective and linear equality constraints, has an analytical solution. The problem can be extended by including position limits or a long-only constraint [23,29,35]. The resulting problem no longer has an analytical solution, but it can be efficiently solved

as a quadratic program (QP) [6, pp. 55–156]. Other constraints and objective terms can be incorporated, including those related to accounting for a previous or initial portfolio [26,34]. Including an initial portfolio allows the method to be used as a trading policy, which can be run periodically to prescribe trades [7].

1.1.2 Portfolio Construction via Convex Optimization

These Markowitz-inspired portfolio construction problems are typically convex and can be efficiently solved [6]. Even complex portfolio construction problems can be specified succinctly in a high-level domain-specific language for convex optimization, such as CVXPY [15], CVX [22], Convex.jl [40], and CVXR [20]. Furthermore, problems with thousands of assets and a risk model with dozens of factors can be solved in well under a second using standard open-source solvers such as ECOS [18], OSQP [38], SCS [32,33], or commercial solvers such as CPLEX [25], MOSEK [31], or GUROBI [24]. Custom implementations of portfolio construction solvers can be far faster, with solve times measured in milliseconds.

Although solver speed is not essential if we are only interested in occasionally rebalancing a handful of portfolios, it can be useful when managing a very large number of individualized accounts. Additionally, fast solvers allow us to quickly run many backtest simulations of a trading algorithm. These simulations allow us to tune hyper-parameters, carry out what-if experiments, and compare different formulations or models on historical or synthesized data.

1.1.3 Nonconvex Portfolio Construction Problems

Some practical constraints and objective terms are not convex. An obvious example is that asset holdings must be in integral numbers of shares. For large portfolios, this constraint is readily handled by simple heuristics, for example by ignoring it in solving the problem, and then rounding the real-valued holdings to the nearest integer values. Other more challenging constraints include limits on the number of assets in the portfolio, or a minimum nonzero trade size [5]. A challenging nonconvex objective term is tax liability, the focus of this paper.

These nonconvex portfolio construction problems can be reformulated as mixed-integer convex optimization problems (which are not convex). They can be solved exactly using a variety of methods and software, such as GLPK [27], CPLEX, MOSEK, and GUROBI. Such solvers are often fast, but for some problem instances can have very long solve times, often hundreds of times more than those associated with similar convex problems. In contrast, solving convex optimization problems is reliably fast.

1.1.4 Convex Approximations

An alternative to solving the nonconvex optimization problem exactly is to employ a heuristic method that finds an approximate solution far faster than it would take to solve the problem exactly. This paper presents one such heuristic method, based on a convex approximation of the original problem. The idea that convex approximations

of nonconvex problems can be used in place of global nonconvex solvers with the same practical performance has been widely noted in other areas [16].

The problem we study in this paper involves the sum of many nonconvex terms that are all similar. Because the sum of a large number of nonconvex functions tends to be ‘more convex’ than the original functions, these problems are often well approximated by convex problems. This intuitive phenomenon was described by Shapley, Folkman, and Starr. Starr applied it to problems in microeconomics involving many agents [37]. [4] provides practical algorithms for solving these problems with performance bounds.

1.1.5 Tax-aware Investment

Our paper is related to literature that develops optimal tax-aware trading strategies, building on the papers by Constantinides [9,10]. The intuition in these papers is that investors may reduce their tax liability by deferring capital gains and realizing losses, which can be used to offset current income or capital gains. This intuition also applies in our setting. Many papers apply well-known numerical methods to solve the tax problem. For example, dynamic programming has been widely applied [11–13]. [19] use a binomial tree and formulate an optimal stopping problem. [14] formulate an optimal tax investment strategy with nonlinear programming. Although Markowitz-based tax-aware portfolio construction is an old idea [34], to our knowledge, ours is the first to develop a *convex* tax optimization problem by relaxing the original nonconvex problem. We focus on the speed and reliability of convex optimization techniques with applications to taxable managed funds [36], the large and rapidly growing tax-loss harvesting industry [8], and security valuation with taxes [21].

1.1.6 Multi-period Portfolio Construction

We focus on the single-period portfolio construction problem, without explicitly planning for future trades. This is in contrast to multi-period portfolio optimization formulations, such as those of [7]. Many tax-aware problems are readily handled by single-period portfolio optimization, such as loss harvesting, tax-neutral portfolio rebalancing, managing inflows and outflows, and optimizing tax-free donations. In fact, we believe repeated single-period portfolio optimization is an excellent approach to certain long-horizon investment problems, such as tracking a (low-turnover) benchmark portfolio while harvesting tax losses. There is some theoretical justification here: [9] shows that for a single asset, with no wash sale rule, the greedy approach of realizing losses and deferring gains is optimal. Repeatedly using single-period, tax-aware portfolio construction does exactly this, but with many assets, while avoiding wash sales. A more practical justification for single-period portfolio optimization is that it is currently standard for such tax-loss harvesting strategies. This paper does not argue for or against using a single-period formulation, but simply gives a reasonable problem formulation and solution method for it.

1.2 Contributions

This paper focuses on incorporating a specific nonconvex term, the tax liability generated by the trades, into an otherwise convex portfolio construction problem. Ignoring this constraint, or using simple ad hoc rounding methods to handle it, does not work well compared to solving the problem exactly with a mixed-integer convex solver. Our contribution is to develop a heuristic method for approximately solving the tax-aware portfolio construction problem that relies on solving two convex optimization problems, making it reliably fast.

2 Tax-Aware Portfolio Construction

This section outlines our notation and describes the tax-aware portfolio optimization problem. We start by describing the trading dynamics and various objective terms.

2.1 Portfolio Holdings and Dynamics

We consider a universe of n assets we are allowed to hold and trade. We let $h_{\text{init}} \in \mathbf{R}^n$ denote the dollar value of our pre-trade holdings of these n assets. We restrict ourselves to long-only portfolios, so $h_{\text{init}} \geq 0$.

Our task is to decide how much of each these assets to buy or sell. We represent this decision by a purchase vector $u \in \mathbf{R}^n$, denominated in dollars. If we purchase asset i , $u_i > 0$; if we sell asset i , $u_i < 0$. Our post-trade holdings are $h \in \mathbf{R}^n$, given by

$$h = h_{\text{init}} + u.$$

This equation ignores transaction costs, which are assumed to be small. (Following convention, we include these transaction costs in our objective function.) We require that the post-trade portfolio is also long-only, making $h \geq 0$. This constraint means we cannot sell more of any asset than we currently hold.

2.1.1 Cash

The cash held in the portfolio is $c_{\text{init}} \in \mathbf{R}$, which we allow to be negative. The post-trade cash balance is

$$c = c_{\text{init}} - \mathbf{1}^T u.$$

We assume the post-trade cash amount must match some desired value c_{des} , which translates to the constraint on u

$$\mathbf{1}^T u = c_{\text{init}} - c_{\text{des}}.$$

The total pre-trade portfolio value, including cash, is $\mathbf{1}^T h_{\text{init}} + c_{\text{init}}$, which we assume is positive. While any value of c_{des} is possible, a common choice is a given

fraction η of the total portfolio value,

$$c_{\text{des}} = \eta(\mathbf{1}^T h_{\text{init}} + c_{\text{init}}). \tag{1}$$

The choice $\eta = 0.01$, for example, means that 1% of the total portfolio value is to be held in cash. The cash balance can be used to handle cash deposits into and withdrawals from the account by adjusting c_{init} by the amount deposited or withdrawn.

2.2 Objective Terms

Here, we describe various objective terms and additional constraints, including the traditional ones: expected return, active risk, and transaction costs. We briefly introduce the tax liability term and provide some of its attributes, reserving a detailed description for Sect. 3.

We note that it is customary to scale the variables and objective terms so that the objective represents an adjusted return. However, to simplify our description of the tax liability function in Sect. 3, we leave the variables and objective in units of dollars.

2.2.1 Risk

The risk of a managed portfolio is typically measured with respect to a benchmark portfolio, such as the S&P 500. This benchmark portfolio is described by a vector $h_b \in \mathbf{R}^n$, scaled so that it has the same market value as our portfolio, *i.e.*, $\mathbf{1}^T h_b = \mathbf{1}^T h_{\text{init}} + c_{\text{init}}$.

The (active) risk is

$$(h - h_b)^T V (h - h_b),$$

where V is the covariance matrix of the asset returns. Our covariance matrix V has the traditional factor model form

$$V = X \Sigma X^T + D,$$

where $X \in \mathbf{R}^{n \times k}$ is the factor exposure matrix, $\Sigma \in \mathbf{R}^{k \times k}$ is the symmetric positive definite factor covariance matrix, and $D \in \mathbf{R}^{n \times n}$ is the diagonal matrix of idiosyncratic variances with $D_{ii} > 0$ [7,23]. The risk can be decomposed into two components, the systematic risk

$$(h - h_b)^T X \Sigma X^T (h - h_b), \tag{2}$$

and the specific risk

$$(h - h_b)^T D (h - h_b) = \sum_{i=1}^n D_{ii} (h_i - h_{b,i})^2. \tag{3}$$

It is common to express active risk in terms of its square root, which has units of dollars. We note for the future use that the specific risk (3) is separable, *i.e.*, a sum of terms each associated with one asset.

2.2.2 Expected Return

Suppose we have a forecast of the return of the n assets, expressed as a vector $\alpha \in \mathbf{R}^n$, where α_i is the expected return of asset i . The expected active return of portfolio h is then $\alpha^T(h - h_b)$, which is measured in dollars. Because a constant offset is immaterial for optimization, we can write the expected return as simply $\alpha^T h$ or even $\alpha^T u$.

2.2.3 Transaction Costs

The transaction cost follows a simple bid-ask spread model:

$$\kappa^T |u|,$$

where $\kappa \in \mathbf{R}_+^n$ is one-half the bid-ask spread, and $|u|$ is the element-wise absolute value of u . For simplicity, we neglect the standard price impact term; this omission is reasonable if we assume our trades are small relative to the total market volume over the trading period. (For larger accounts, a price impact term can be included; see [7].)

2.2.4 Tax Liability

We let $L : \mathbf{R}^n \rightarrow \mathbf{R}$ denote the tax liability function, where $L(u)$ is the immediate tax liability incurred by the trades u due to realizing capital gains. We will describe $L(u)$, which derives from the history of previous transactions in the assets, in detail in Sect. 3; for now, we simply note some of its attributes. First, it is separable across the assets, *i.e.*, it has the form

$$L(u) = \sum_{i=1}^n L_i(u_i),$$

where $L_i(u_i)$ is the tax liability for asset i incurred by trading. There is no immediate tax liability when buying an asset, making $L_i(u_i) = 0$ for $u_i \geq 0$. For $u_i < 0$, *i.e.*, selling the asset, $L_i(u_i)$ is a convex piecewise linear function. While L_i is convex for $u_i < 0$, it is not convex over the whole interval, which includes buying ($u_i > 0$) and selling ($u_i < 0$). The total tax liability function $L(u)$ is not convex, but it becomes convex if we restrict the sign of u_i , *i.e.*, we specify whether we are buying or selling each asset.

2.2.5 Constraints

We have already mentioned several constraints, for example that $h \geq 0$ (the portfolio is long-only) and $\mathbf{1}^T u = c_{\text{init}} - c_{\text{des}}$ (the post-trade cash matches a desired value). We

also allow for additional convex constraints on the trade list and post-trade holdings. We represent these as $u \in \mathcal{U}$ and $h \in \mathcal{H}$. These could include, for example, limits on the holdings of a particular asset, or limits on the exposure of our portfolio to a certain factor. For concreteness, we assume \mathcal{H} and \mathcal{U} are polyhedral, *i.e.*, described by a finite set of linear equality and inequality constraints, although our proposed method also applies more generally.

2.3 Tax-Aware Portfolio Construction

2.3.1 Tax-Aware Utility Function

We assemble our objective terms into a single utility function of u and h ,

$$U(h, u) = \alpha^T u - \gamma_{\text{risk}}(h - h_b)^T V(h - h_b) - \gamma_{\text{tc}} \kappa^T |u| - \gamma_{\text{tax}} L(u). \quad (4)$$

where γ_{risk} , γ_{tc} , and γ_{tax} are nonnegative trade-off parameters. The first two terms constitute the traditional risk-adjusted return used in Markowitz portfolio construction. The third term is transaction cost, a widely used addition to the traditional Markowitz utility, with the parameter γ_{tc} used to control turnover. The last term accounts for the tax liability of the trades.

2.3.2 Tax-Aware Portfolio Construction Problem

Our problem is to maximize utility subject to constraints, *i.e.*,

$$\begin{aligned} & \text{maximize } \alpha^T u - \gamma_{\text{risk}}(h - h_b)^T V(h - h_b) - \gamma_{\text{tc}} \kappa^T |u| - \gamma_{\text{tax}} L(u) \\ & \text{subject to } h = h_{\text{init}} + u, \quad \mathbf{1}^T u = c_{\text{init}} - c_{\text{des}} \\ & \quad u \in \mathcal{U}, \quad h \in \mathcal{H}, \end{aligned} \quad (5)$$

with decision variables u and h . The problem data are α , h_b , V , κ , h_{init} , c_{des} , c_{init} , the function L (described in Sect. 3), the constraint sets \mathcal{U} and \mathcal{H} , and the trade-off parameters γ_{risk} , γ_{tc} , and γ_{tax} . We refer to the problem (5) as the *tax-aware Markowitz problem*, or TAM problem, and we denote its optimal value as U^* .

2.3.3 Nonconvexity

The constraints in the TAM problem (5) are convex, as are all terms in the objective with the exception of the tax liability term. Unfortunately, that term renders the TAM problem (5) nonconvex, which makes it difficult to solve (exactly) in general. We note, however, that the problem becomes convex when we specify the sign of the trade list u , *i.e.*, if we specify for each asset whether we are to sell it ($u_i \leq 0$) or buy it ($u_i \geq 0$).

The TAM problem can be formulated as a mixed-integer quadratic program (MIQP), which can be solved using various methods. It is well known that in practice, these methods can often solve problems reasonably quickly, but in many other cases, the solution times can be extremely long. The main contribution of this paper is a method

for approximately solving the TAM problem, which involves solving only two convex optimization problems. As a result, our method is always very fast and never involves the very long solution time that can be observed with MIQP solvers. As we will see in Sect. 6, for realistic instances of the TAM problem, our method delivers near identical performance as a globally optimal solution.

3 Tax Liability

This section describes the tax liability function.

3.1 Tax Lots and Capital Gains

3.1.1 Tax Lots

For each asset, the pre-trade holdings are composed of zero or more tax lots. Each tax lot has several attributes associated with it: its quantity of shares, acquisition date, and cost basis (the price per share at which the shares were acquired).

We let q_{ij} denote the quantity of shares in the j th lot of asset i . The total number of shares of asset i held is $\sum_j q_{ij}$. We have $h_{\text{init},i} = p_i \sum_j q_{ij}$, where p_i is the current price of asset i . We let b_{ij} denote the cost basis (in dollars per share) of the j th lot of asset i .

3.1.2 Selling Shares

When shares of asset i are sold, *i.e.*, we have $u_i < 0$, we must specify which tax lots from which to take the shares. Let s_{ij} denote the dollar value of shares sold from the j th lot of asset i , with $0 \leq s_{ij} \leq q_{ij} p_i$, where $q_{ij} p_i$ is the dollar value of the j th lot of asset i . The total dollar value of asset i sold is then $\sum_j s_{ij}$, which must be equal to $-u_i$.

When we sell s_{ij} dollars from lot j of asset i , we incur a capital gain, which is the difference of our proceeds and our cost basis for those shares, *i.e.*, $(1 - b_{ij}/p_i)s_{ij}$. We refer to this quantity as the gain; when it is negative, we refer to it as the loss.

3.1.3 Long-Term and Short-Term Gains

A tax lot is *long term* if the acquisition date is more than 1 year before the trade date, and the lot is *short term* otherwise. Gains from long-term and short-term lots are taxed at two different positive rates, ρ_{lt} and ρ_{st} , respectively, with $\rho_{\text{lt}} \leq \rho_{\text{st}}$. The tax liability for selling dollar value s_{ij} from the j th lot of asset i is $\rho_{\text{lt}}(1 - b_{ij}/p_i)s_{ij}$ if lot j is long term, and $\rho_{\text{st}}(1 - b_{ij}/p_i)s_{ij}$ if lot j is short term.

The total tax liability from selling all assets is

$$\sum_{i,j} \rho_{ij}(1 - b_{ij}/p_i)s_{ij} = \sum_{i,j} T_{ij}s_{ij},$$

where the tax rates ρ_{ij} are given by:

$$\rho_{ij} = \begin{cases} \rho_{lt} & \text{lot } j \text{ of asset } i \text{ is long term} \\ \rho_{st} & \text{lot } j \text{ of asset } i \text{ is short term.} \end{cases}$$

We refer to $T_{ij} = \rho_{ij}(1 - b_{ij}/p_i)$ as the tax rate for lot j of asset i . This is the dollar tax liability generated per dollar sold of the lot. It is positive if the current asset price exceeds the lot basis, *i.e.*, the lot is held at a gain, and is negative if the lot is held at a loss.

3.2 Tax Liability Function

Suppose that for asset i we have $u_i < 0$, *i.e.*, we are selling $-u_i$ dollars of asset i , which translates to $-u_i/p_i$ shares. We can solve the problem of allocating the sale across lots in order to minimize the tax liability incurred. We define

$$L_i(u_i) = \min_{s_{ij}} \left\{ \sum_j T_{ij}s_{ij} \mid \sum_j s_{ij} = -u_i, 0 \leq s_{ij} \leq q_{ij}p_i \right\},$$

which is the smallest tax gain achievable to carry out this sale. We define $L_i(u_i) = +\infty$ for $-u_i < p_i \sum_j q_{ij}$, *i.e.*, if we ask to sell more shares of the asset than we hold. We also define $L_i(u_i) = 0$ for $u_i \geq 0$, *i.e.*, we are buying shares instead of selling. These properties hold because purchasing additional shares incurs no immediate tax liability.

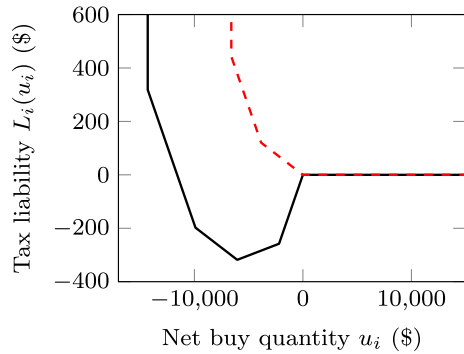
3.2.1 Least-Tax-First-Out Lot Policy

For a given value of $u_i < 0$, it is easy to determine optimal values of s_{ij} ; it is a convex optimization problem with an analytical solution. We simply sort the values of T_{ij} from least (most negative) to greatest, breaking ties arbitrarily. Then we sell shares from lots in this order. For example, we start by selling shares from the lot with the smallest (or most negative) value of T_{ij} , which is the term-adjusted tax liability rate. If we need to sell more shares than that, we go to the lot with second smallest value, and so on. This greedy approach is optimal, *i.e.*, it minimizes the tax liability when selling $-u_i$ dollars of asset i . We refer to this approach of choosing lots from which to sell shares as *least-tax-first-out* (LTFO). It takes into account whether the lots are long term or short term. If all lots are short term, this scheme corresponds to the well-known *highest-basis-first-out* (HIFO) method.

The tax liability function L_i is continuous and piecewise affine. If none of the lots are at a loss, L_i is convex and nonnegative. If at least one lot is held at a loss, then L_i takes on negative values and is *not* convex. When the domain of L_i is restricted to either $u_i \leq 0$ or $u_i \geq 0$, the resulting function is convex.

Figure 1 shows two different tax liability functions. The dashed red curve shows the tax liability function for an asset which we hold in two lots, both at a gain (*i.e.*, with current price greater than basis). The solid black curve shows the tax liability

Fig. 1 Tax liability functions L_i for two assets. The solid black curve is for an asset with four lots with two held at a loss and two held at a gain. The dashed red curve is for an asset with two lots, both held at a gain



function of a different asset, which we hold in four lots, two at a loss (*i.e.*, the basis is greater than the current price), and two at a gain. Each linear segment corresponds to a tax lot, with the slope given by the tax rate of the lot, and width given by the total value of the lot.

4 Convex Relaxation

The first step in developing our convex-optimization-based heuristic for approximately solving the TAM problem is to form a convex relaxation that approximates the problem.

4.1 Convex Envelope of a Function

We review a standard concept, the *convex envelope* of a function $f : \mathbf{R} \rightarrow \mathbf{R}$, denoted f^{**} . It is defined as

$$f^{**}(x) = \inf\{\theta f(v) + (1 - \theta)f(w) \mid \theta \in [0, 1], x = \theta v + (1 - \theta)w\}. \tag{6}$$

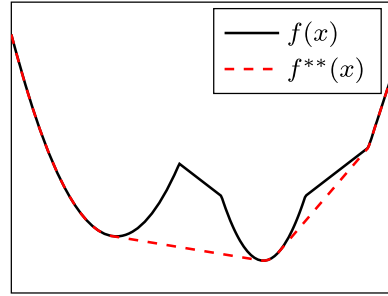
The infimum is over θ , v , and w . The convex envelope function f^{**} is convex, and it satisfies $f^{**}(x) \leq f(x)$ for all x , *i.e.*, it is a global underestimator of f . If f is convex, then f^{**} is equal to f .

The convex envelope can be defined several other equivalent ways. For example, f^{**} is the greatest convex function that is a global underestimator of f . It is also the (Fenchel) conjugate of the conjugate of f , *i.e.*, $(f^*)^*$, where the superscript $*$ is the traditional notation for the conjugate function. (This explains why we denote the convex envelope of f as f^{**} .) An example is shown in Fig. 2.

If $f(x)$ is convex when restricted to $x \leq 0$ and also when restricted to $x \geq 0$, we can require that $v \geq 0$ and $w \leq 0$ in (6), *i.e.*, we can define the convex envelope as

$$f^{**}(x) = \inf\{\theta f(v) + (1 - \theta)f(w) \mid \theta \in [0, 1], x = \theta v + (1 - \theta)w, v \geq 0, w \leq 0\}. \tag{7}$$

Fig. 2 A nonconvex function f (solid black) and its convex envelope f^{**} (dashed red)



4.2 Convex Relaxation with Borrowed Curvature

In this section, we describe a convex relaxation of the TAM problem (5) using the convex envelope. We first eliminate the post-trade holdings variable h to express the problem in terms of the trade list u , with an objective that is the sum of a separable function and one that is not separable. This form is:

$$\begin{aligned} &\text{maximize } -f_0(u) - \sum_{i=1}^n f_i(u_i) \\ &\text{subject to } u \in \tilde{\mathcal{U}} \end{aligned} \tag{8}$$

with variable $u \in \mathbf{R}^n$, where the constraint set is

$$\tilde{\mathcal{U}} = \{u \in \mathcal{U} \mid h_{\text{init}} + u \in \mathcal{H}, \mathbf{1}^T u = c_{\text{init}} - c_{\text{des}}\}.$$

The constraint set $\tilde{\mathcal{U}}$ includes the original constraint $u \in \mathcal{U}$ as well as the holdings constraint $h \in \mathcal{H}$ and the post-trade cash constraint, and is convex. The nonseparable part of the objective function is

$$f_0(u) = \gamma_{\text{risk}}(h_{\text{init}} - h_b + u)^T X \Sigma X^T (h_{\text{init}} - h_b + u),$$

which is the systematic component of risk (2), and is convex. The separable part corresponding to asset i is

$$f_i(u_i) = -\alpha_i u_i + \gamma_{\text{risk}} D_{ii} (h_{\text{init},i} - h_{b,i} + u_i)^2 + \gamma_{\text{tc}} \kappa_i |u_i| + \gamma_{\text{tax}} L_i(u_i), \tag{9}$$

which includes contributions from the expected return, specific risk (3), transaction cost, and tax liability.

The functions f_i are piecewise quadratic and nonconvex in general, but are convex when $u_i \leq 0$ or $u_i \geq 0$. The problem (8), which is equivalent to the original TAM problem (5), is not convex because the functions f_i are not convex. However, if we fix the sign of each u_i , the problem (8) becomes convex, and therefore easy to solve. (In fact, it suffices to fix the sign of u_i for each asset where we hold at least one lot at a loss; the other f_i are convex.)

4.2.1 Relaxed TAM Problem with Borrowed Curvature

We can now obtain a convex relaxation of the TAM problem by replacing f_i with f_i^{**} , to obtain

$$\begin{aligned} &\text{maximize } -f_0(u) - \sum_{i=1}^n f_i^{**}(u_i) \\ &\text{subject to } u \in \tilde{\mathcal{U}}. \end{aligned} \tag{10}$$

This is a convex problem, which can be formulated as a second-order cone problem (SOCP) [6, §4.4.2]. The convex envelopes f_i^{**} are convex and also piecewise quadratic. Fig. 3 plots f_i and f_i^{**} .

The objective of the relaxation (10) is an upper bound on the objective of the original TAM problem. It follows that its optimal objective value U_{relax}^* is an upper bound on the optimal value of the original TAM problem. It follows that its optimal objective value U_{relax}^* is an upper bound on the optimal value of the original TAM problem, *i.e.*,

$$U^* \leq U_{\text{relax}}^*.$$

The gap $U_{\text{relax}}^* - U^*$ can be bounded in terms of k , the number of factors in the risk model, and the distances between the separable functions f_i and their convex envelopes f_i^{**} . This is an application of the Shapley–Folkman Lemma [3,39].

4.2.2 TAM Problem with Approximate Tax Liability

By re-introducing the post-trade holdings variable, the relaxed problem (10) can be written as

$$\begin{aligned} &\text{maximize } \alpha^T u - \gamma_{\text{risk}}(h - h_b)^T V(h - h_b) - \gamma_{\text{tc}} \kappa^T |u| - \gamma_{\text{tax}} \hat{L}(u) \\ &\text{subject to } h = h_{\text{init}} + u, \quad \mathbf{1}^T u = c_{\text{init}} - c_{\text{des}} \\ &\quad u \in \mathcal{U}, \quad h \in \mathcal{H}, \end{aligned} \tag{11}$$

with decision variables u and h . This is the TAM problem with the tax liability functions L replaced by *approximate* tax liability function \hat{L} , defined as

$$\hat{L}(u) = \sum_{i=1}^n \hat{L}_i(u_i),$$

where $\hat{L}_i = L_i + (f_i^{**} - f_i)/\gamma_{\text{tax}}$. Problem (11) can be solved exactly using convex optimization, even though the functions \hat{L}_i are not convex. This is possible because the nonconvex function \hat{L} borrows curvature from the other separable objective terms, resulting in an objective function that is concave (see Fig. 3.)

5 Approximate Solution Methods

This section describes heuristic solution methods for the TAM problem. The methods involve solving two convex optimization problems in two stages.

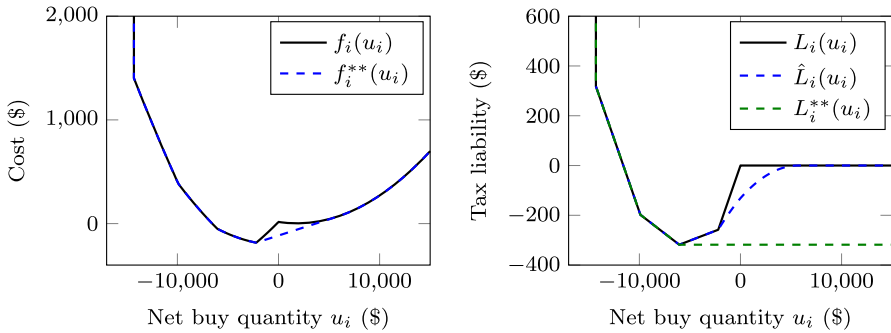


Fig. 3 *Left.* $f_i(u_i)$ (solid black) and its convex envelope $f_i^{**}(u_i)$ (blue dashed). *Right.* Nonconvex tax liability function $L_i(u_i)$ (solid black), its convex envelope $L_i^{**}(u_i)$ (dashed green line), and the approximation used in our relaxation $\hat{L}_i(u_i)$ (dashed blue). Even though $\hat{L}_i(u_i)$ is nonconvex, we can still solve the problem globally and efficiently.

1. *Guess the vector z of signs of an optimal u .* This is done by solving the relaxation (10) of the TAM problem (which in addition provides an upper bound on U^*).
2. *Solve TAM with these sign constraints.* Add the sign constraints $z_i u_i \geq 0, i = 1, \dots, n$ to the TAM problem and solve. With these constraints, the TAM problem is a convex QP and can be efficiently solved.

There are several choices for step 1, which we describe below. We only need to specify the sign of u_i for assets in which we hold at least one lot at a loss.

5.1 Methods for Guessing the Sign

For step 1, we solve or the relaxation (10). There are several choices for guessing the signs of u_i from the solution of one of this relaxation. The most obvious method is to use the sign of the solution of the relaxation, *i.e.*, $z = \text{sign}(u_{\text{relax}}^*)$.

A less obvious method is a random choice of the signs with probabilities taken from the solution of the relaxation. For each i , we obtain the values θ_i in the convex envelope definition (7) for f_i . We then set $z_i = 1$ with probability θ_i , and $z_i = -1$ with probability $1 - \theta_i$. (This is done independently for each i .) Thus, we use the values in the convex envelope as probabilities on whether we buy or sell each asset. This method can be used to generate multiple candidate sign vectors, and we can compare the objectives after step 2 and use the one with the largest objective.

In many numerical experiments, we found that the method that performs best is to solve the relaxation (10), and then use the randomized method to guess a set of signs. (This is despite the fact that the simple rounding method is guaranteed to produce feasible sign constraints for the TAM problem, and the randomized method is not.) We have also found that generating multiple sets of candidate signs does not substantially improve the results. This method requires two convex optimization solves: one to solve the relaxation (an SOCP), and one to solve the original TAM problem with the sign of u fixed (a QP).

6 Numerical Examples

We demonstrate these methods by simulating a tax-loss harvesting strategy, in which we solve the TAM problem once a month to generate the trade list. First, we show a 6-year backtest of such a strategy. Then, we use this backtest (and others like it) to generate realistic instances of the TAM problem, which we use to evaluate the methods of Sect. 5.

6.1 Benchmark and Data

All of our simulations use the S&P 500 as the benchmark, with data over the period 2002 to 2019. Our universe includes all assets that were in the S&P 500 at any point over that time interval, which gives $n = 998$. We included a constraint that we only purchase shares of current S&P 500 constituents. This prevents us from purchasing assets that, at the time of the simulated trade, have never been in the benchmark. It also means we don't increase our holdings of former S&P 500 constituents (but we also do not require them to be immediately sold). If any asset is delisted, we liquidate the asset immediately, incurring the associated tax liability.

We take $\alpha = 0$, *i.e.*, we do not have any views on the active returns, so our goal is to simply track the benchmark portfolio while minimizing tax liability. Our risk model parameters Σ , X , and D are from the Barra US Equity model [30], which uses $k = 72$ factors. Our cash target c_{des} is given by (1) with $\eta = 0.005$, *i.e.*, we hold 50 basis points in cash after each trade. We use tax rates $\rho_{\text{lt}} = 0.238$ and $\rho_{\text{st}} = 0.408$, which reflect the current highest marginal tax rates in the United States for long-term and short-term capital gains, respectively. We used the conservative value $\kappa_i = 0.0005$ for all transaction costs, *i.e.*, the bid-ask spread is 10 basis points for all assets. The parameter γ_{risk} was scaled with the account value, so that $\gamma_{\text{risk}} = \tilde{\gamma}_{\text{risk}}(\mathbf{1}^T h_{\text{init}} + c_{\text{init}})$, with $\tilde{\gamma}_{\text{risk}} = 200$. The other trade-off parameters were $\gamma_{\text{tc}} = 1$ and $\gamma_{\text{tax}} = 1$.

6.2 Backtests

Our dataset consists of 204 months over a 17-year period from August 2002 through August 2019. We use this dataset to carry out 12 different, staggered 6-year-long backtests. The first one starts in August 2002 and ends in July 2008; the last one starts in August 2013 and ends in July 2019. In these backtests, monthly trading means we trade on the first business day more than 31 calendar days after the last trade. For each trade, the initial cash amount c_{init} is adjusted for the realized transaction cost $\kappa^T |u|$ of the last trade, as well as cash inflows due to dividends and other corporate actions. In the backtests, we round the trade lists to an integer number of shares. Each backtest starts with a portfolio of \$1M in cash.

Each month, the trade list is determined by solving the TAM problem using one of two methods:

- *Heuristic* We solve the relaxation (10) and use the randomized rounding method.
- *Mixed-integer method* We use the mixed-integer mode of CPLEX (version 12.9) to solve the TAM problem directly, with a time limit of 300 seconds.

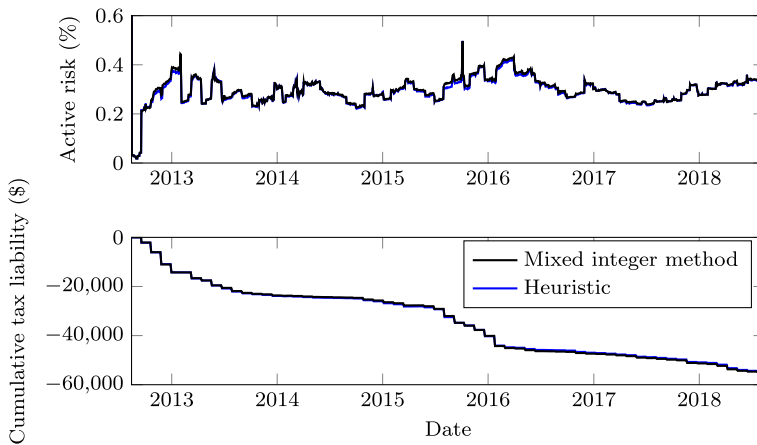


Fig. 4 The active risk (top) and cumulative tax liability (bottom) of a backtest for both solution methods

For the heuristic, we used CPLEX (as a QP/SOCP solver) to solve the convex relaxations and to generate the final trade list.

6.2.1 Example Backtest

Figure 4 shows the results of one of our backtests, initiated in August 2013. The top plot shows the active risk, and the bottom plot shows the cumulative tax liability, which is the net realized gain, accounting for long- and short-term tax rates. (This quantity is negative, meaning we are realizing a net loss). Here, we use the conventional definition of active risk, which is the square root of the definition given in Sect. 2 and is scaled down by the account value.

These results show that a tax-aware trading scheme can indeed track a benchmark while simultaneously realizing capital losses. It is interesting to note that losses are harvested even during bull markets. The rate of tax-loss harvesting decreases with the life of the fund, since more lots are held at a gain. We note that the two methods have nearly indistinguishable performance in the backtest. The backtests we ran over other time windows had similar outcomes.

6.3 Detailed Comparison of Solution Methods

This section compares the performance of both solution methods. We use the data from 12 backtests, each 6 years long, giving a total of 744 instances of the TAM problem. (We exclude the initial trade, in which the account holds only cash.) For these problem instances, we compute the utility achieved by the heuristic method, denoted $U_{\text{relax,round}}$, the utility of the mixed-integer method, denoted U_{mip} , as well as the upper bound U_{relax}^* . To make the utility (4) comparable across the problem instances, we divide it by the account value $\mathbf{1}^T h_{\text{init}} + c_{\text{init}}$. (This number is the monthly after-tax expected return adjusted for risk and transaction costs and is measured in percent or

Table 1 Comparison of the solution methods. The first two columns show the number of problem instances (out of 744 total) for which the achieved utility U of each method exceeds the mixed-integer method's utility U_{mip} (first column) or matches the upper bound U_{relax}^* exactly (second column)

	Improved vs MIP $U \geq U_{\text{mip}}$ (count)	Solved $U = U_{\text{relax}}^*$ (count)	Mean subopt. gap $U_{\text{relax}}^* - U$ (bp)	Mean MIP gap $U_{\text{mip}}^* - U$ (bp)
Heuristic	706	678	0.02	-0.01
MIP (300s)	744	646	0.03	0

Equality here is to within 0.05 bp, the approximate solver tolerance. The last two columns give the average difference between the achieved utility U and U_{mip} (third column) or U_{relax}^* (fourth column)

basis points.) The mean optimal utility U^* across these 744 problems is 2 basis points, and the standard deviation is 55 basis points.

6.3.1 Evaluation of the Heuristic

The results are shown in Table 1. We see that for 678 of the 744 instances, $U_{\text{relax,round}} = U_{\text{relax}}^*$, *i.e.*, the heuristic solves the TAM problem and provides a certificate of optimality, to within the solver numerical accuracy, which is around 0.05 bp. We also see that the (mean) differences in utility obtained by the heuristic and the bound are very close, *i.e.*, within fractions of a basis point. To summarize, the heuristic produces an optimal trade list (and certificate of optimality) for the vast majority of the problem instances. We note that for all 744 problem instances, the heuristic is never suboptimal by more than a few basis points.

We now compare the heuristic to the mixed-integer method. Note that if CPLEX solves the mixed-integer problem within the 300 second time limit, we have $U_{\text{mip}} = U^*$. The mixed-integer method times out (and therefore, does not necessarily globally solve the problem) in 179 of the 744 cases. Among the 565 instances that the mixed-integer method solves (within 300 seconds), in 549 instances the heuristic method also solves the problem to within 0.05 basis points, and is never more than 0.3 basis points suboptimal in the remaining 16 cases. For the 179 cases in which the mixed-integer method fails to solve the problem, we observe that the heuristic achieves a normalized utility within 0.05 basis points of the mixed-integer method in 89 cases, and outperforms it by more than 0.05 basis points in 68 cases (by up to 2 basis points). In only 22 of the 179 instances did the heuristic underperform the mixed-integer method by more than 0.05 basis points, and never underperformed by more than 2 basis points.

6.3.2 Algorithm Run Times

Figure 5 shows the algorithm run times for the heuristic and the mixed-integer method on a scatter plot. All of the points are below the dashed black line, which indicates that the heuristic method was faster in all cases. Out of the 744 problem instances, 179 took 300 s using the mixed-integer method, which was the maximum time allowed.

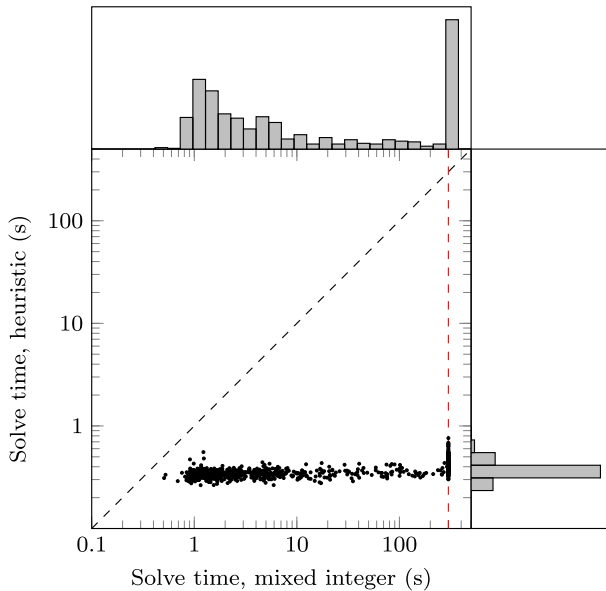


Fig. 5 The algorithm run times of the 720 problem instances using the relax-and-round heuristic and the mixed-integer solution, with each problem instance shown as a single dot. The dashed red line shows the maximum allowed time of the mixed-integer solver

7 Conclusion

We formulate tax-aware portfolio construction as a nonconvex optimization problem, and we present a heuristic for this problem based on convex optimization. This method is reliably fast: for problems with several hundred assets and several dozen factors, it takes less than a second. We compare our heuristic against the standard, mixed-integer quadratic programming formulation, solved using CPLEX, on realistic problem instances. When the mixed-integer method is limited to five minute run times, we find that our heuristic outperforms it more often than not, despite being several hundred times faster. This speed is not necessary for monthly (or even daily) trading, but is useful for backtesting and Monte Carlo simulation, possibly over hundreds of thousands of individualized accounts. Our method also produces a bound on the optimal value. For realistic data, the bound is usually tight enough that it certifies that the heuristic solved the problem globally. In the future work, we will extend this method to other nonconvex terms that are often present in practical portfolio optimization problems.

Acknowledgements We would like to thank Emmanuel Candès for useful discussions and feedback. We would also like to thank Eric Kisslinger for identifying an important error in an early version of the software.

References

1. Atra, R., Pae, Y.: Likely benefits from FIFO accounting. *J. Financ. Plan.* **27**(1), 54 (2014)

2. Berkin, A., Ye, J.: Tax management, loss harvesting, and FIFO accounting. *Financ. Anal. J.* **59**(4), 91–102 (2003)
3. Bertsekas, D.P.: *Constrained optimization and Lagrange multiplier methods*. Academic Press (1982)
4. Bertsekas, D.P.: *Nonlinear programming*, 3rd edn. Athena Scientific (1999)
5. Bertsimas, D., Darnell, C., Soucy, R.: Portfolio construction through mixed-integer programming at Grantham, Mayo Van Otterloo and Company. *Interfaces* **29**(1), 49–66 (1999)
6. Boyd, S., Vandenberghe, L.: *Convex Optimization*. Cambridge University Press, Cambridge (2004)
7. Boyd, S., Busseti, E., Diamond, S., Kahn, R., Nystrup, P., Speth, J.: Multi-period trading via convex optimization. *Found. Trends Optim.* **3**(1), 1–76 (2017)
8. Chaudhuri, S.E., Burnham, T.C., Lo, A.W.: An empirical evaluation of tax-loss-harvesting alpha. *Financ. Anal. J.* **76**(3), 99–108 (2020)
9. Constantinides, G.: Capital market equilibrium with personal tax. *Econometrica* **51**(3), 611–636 (1983)
10. Constantinides, G.: Optimal stock trading with personal taxes. *J. Financ. Econ.* **13**(1), 65–89 (1984)
11. Dammon, R., Spatt, C.: The optimal trading and pricing of securities with asymmetric capital gains taxes and transaction costs. *Rev. Financ. Stud.* **9**(3), 921–952 (1996)
12. Dammon, R.M., Spatt, C.S., Zhang, H.H.: Optimal consumption and investment with capital gains taxes. *Rev. Financ. Stud.* **14**(3), 583–616 (2001)
13. Dammon, R., Spatt, C., Zhang, H.: Optimal asset location and allocation with taxable and tax-deferred investing. *J. Finance* **59**(3), 999–1037 (2004)
14. DeMiguel, V., Uppal, R.: Portfolio investment with the exact tax basis via nonlinear programming. *Manage. Sci.* **51**(2), 277–290 (2005)
15. Diamond, S., Boyd, S.: CVXPY: A Python-embedded modeling language for convex optimization. *J. Mach. Learn. Res.* **17**(83), 1–5 (2016)
16. Diamond, S., Takapoui, R., Boyd, S.: A general system for heuristic minimization of convex functions over non-convex sets. *Optim. Methods Softw.* **33**(1), 165–193 (2018)
17. Dickson, J., Shoven, J., Sialm, C.: Tax externalities of equity mutual funds. *Natl. Tax J.* **53**(3), 607–628 (2000)
18. Domahidi, A., Chu, E., Boyd, S.: “ECOS: An SOCP solver for embedded systems”. In: *European Control Conference*, pp. 3071–3076 (2013)
19. Dybvig, P., Koo, H.: *Investment with Taxes*. Washington University in Saint Louis, Tech. rep (1996)
20. Fu, A., Narasimhan, B., Boyd, S.: CVXR: an R package for disciplined convex optimization. *J. Stat. Softw.* **94**(14), 1–34 (2020)
21. Gallmeyer, M., Srivastava, S.: Arbitrage and the tax code. *Math. Financ. Econ.* **4**(3), 183–221 (2011)
22. Grant, M., Boyd, S.: CVX: Matlab Software for Disciplined Convex Programming, version 2.1. <http://cvxr.com/cvx> (Mar. 2014)
23. Grinold, R., Kahn, R.: *Active Portfolio Management*, 2nd edn. McGraw- Hill, New York (1999)
24. Gurobi Optimization LLC: *Gurobi Optimizer Reference Manual*. <http://www.gurobi.com> (2020)
25. IBM Corporation: CPLEX. https://www.ibm.com/support/knowledgecenter/SSSA5P_12.9.0/ilog.odms.studio.help/Optimization_Studio/topics/COS_home.html. Version 12.9 (2019)
26. Lobo, M., Fazel, M., Boyd, S.: Portfolio optimization with linear and fixed transaction costs. *Ann. Oper. Res.* **152**(1), 341–365 (2007)
27. Makhorin, A.: GNU Linear Programming Kit. <https://www.gnu.org/software/glpk/>. GNU Project (2016)
28. Markowitz, H.: Portfolio Selection. *J. Finance* **7**(1), 77–91 (1952)
29. Markowitz, H.: *The Optimization of a Quadratic Function Subject to Linear Constraints*. RAND Corporation, Tech. rep (1955)
30. Menchero, J., Orr, D., Wang, J.: *The Barra US equity model (USE4), methodology notes*. English, MSCI (May 2011)
31. MOSEK ApS: *The MOSEK optimization toolbox for MATLAB manual, version 9.0*. <http://docs.mosek.com/9.0/toolbox/index.html> (2019)
32. O’Donoghue, B., Chu, E., Parikh, N., Boyd, S.: Conic optimization via operator splitting and homogeneous self-dual embedding. *J. Optim. Theory Appl.* **169**(3), 1042–1068 (2016)
33. O’Donoghue, B., Chu, E., Parikh, N., Boyd, S.: SCS: Splitting Conic Solver, version 2.1.2. <https://github.com/cvxgrp/scs> (Nov. 2019)
34. Pogue, G.: An extension of the Markowitz portfolio selection model to include variable transactions’ costs, short sales, leverage policies and taxes. *J. Finance* **25**(5), 1005–1027 (1970)
35. Sharpe, W.: A simplified model for portfolio analysis. *Manag. Sci.* **9**(2), 277–293 (1963)

36. Sialm, C., Zhang, H.: Tax-efficient asset management: evidence from equity mutual funds. *J. Finance* **75**(2), 735–777 (2020)
37. Starr, R.: “Quasi-equilibria in markets with non-convex preferences”. In: *Econometrica: J. Econ. Soc.*, pp. 25–38 (1969)
38. Stellato, B., Banjac, G., Goulart, P., Bemporad, A., Boyd, S.: OSQP: An Operator Splitting Solver for Quadratic Programs. To appear, *Mathematical Programming Computation* (2020)
39. Udell, M., Boyd, S.: Bounding duality gap for separable problems with linear constraints. *Comput. Optim. Appl.* **64**(2), 355–378 (2016)
40. Udell, M., Mohan, K., Zeng, D., Hong, J., Diamond, S., Boyd, S.: “Convex optimization in Julia”. In: *Workshop for High Performance Technical Computing in Dynamic Languages*. IEEE, pp. 18–28 (2014)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.