

# Embedded Code Generation With CVXPY

Maximilian Schaller<sup>1b</sup>, Goran Banjac<sup>1b</sup>, Steven Diamond, *Member, IEEE*, Akshay Agrawal, Bartolomeo Stellato<sup>1b</sup>, *Member, IEEE*, and Stephen Boyd<sup>1b</sup>, *Fellow, IEEE*

**Abstract**—We introduce CVXPYgen, a tool for generating custom C code, suitable for embedded applications, that solves a parameterized class of convex optimization problems. CVXPYgen is based on CVXPY, a Python-embedded domain-specific language that supports a natural syntax (that follows the mathematical description) for specifying convex optimization problems. Along with the C implementation of a custom solver, CVXPYgen creates a Python wrapper for prototyping and desktop (non-embedded) applications. We give two examples, position control of a quadcopter and back-testing a portfolio optimization model. CVXPYgen outperforms a state-of-the-art code generation tool in terms of problem size it can handle, binary code size, and solve times. CVXPYgen and the generated solvers are open-source.

**Index Terms**—Computational methods, embedded systems, optimization.

## I. INTRODUCTION

CONVEX optimization is used in many domains, including signal and image processing [1], [2], control [3], [4], and finance [5], [6], to mention just a few. A (parameterized) convex optimization problem can be written as

$$\begin{aligned} & \text{minimize } f_0(x, \theta) \\ & \text{subject to } f_i(x, \theta) \leq 0, \quad i = 1, \dots, p \\ & \quad \quad \quad g_j(x, \theta) = 0, \quad j = 1, \dots, r, \end{aligned} \quad (1)$$

where  $x \in \mathbf{R}^n$  is the optimization variable,  $f_0$  is the objective function to be minimized,  $f_1, \dots, f_p$  are the inequality constraint functions, and  $g_1, \dots, g_r$  are the equality constraint functions.

Manuscript received March 21, 2022; accepted April 21, 2022. Date of publication May 6, 2022; date of current version May 16, 2022. This work was supported in part by the European Research Council (ERC) through the European Union's Horizon 2020 Research and Innovation Programme grant agreement under Grant 787845; in part by the Stanford's SystemX; and in part by the AI Chip Center for Emerging Smart Systems (ACCESS). Recommended by Senior Editor F. Dabbene. (*Corresponding author: Maximilian Schaller.*)

Maximilian Schaller and Goran Banjac are with the Department of Information Technology and Electrical Engineering, ETH Zürich, 8092 Zürich, Switzerland (e-mail: mschaller@ethz.ch; gbanjac@ethz.ch).

Steven Diamond is with Gridmatic, Campbell, CA 95008 USA (e-mail: steven@gridmatic.com).

Akshay Agrawal and Stephen Boyd are with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA (e-mail: akshayka@stanford.edu; boyd@stanford.edu).

Bartolomeo Stellato is with the Department of Operations Research and Financial Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: bstellato@princeton.edu).

Digital Object Identifier 10.1109/LCSYS.2022.3173209

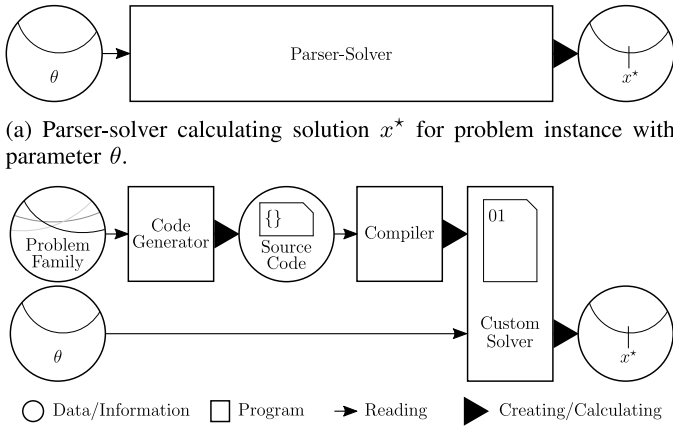
We require that  $f_0, \dots, f_p$  are convex functions, and  $g_1, \dots, g_r$  are affine functions [7]. The parameter  $\theta \in \mathbf{R}^d$  specifies data that can change, but is constant and given when we solve an instance of the problem. We refer to the parameterized problem (1) as a *problem family*; when we specify a fixed value of  $\theta$ , we refer to it as a *problem instance*. We let  $x^*$  denote an optimal point for the problem (1), assuming it exists.

The problem family can be specified using a domain-specific language (DSL) for convex optimization. Such systems allow the user to specify the functions  $f_i$  and  $g_j$  in a simple format that closely follows the mathematical description of the problem. Examples include YALMIP [8] and CVX [9] (in MATLAB), CVXPY [10] (in Python), Convex.jl [11] and JuMP [12] (in Julia), and CVXR [13] (in R). We focus on CVXPY, which provides an efficient way of dealing with parameters, specifying problem families, not just problem instances.

DSLs parse the problem description and translate (or canonicalize) it to an equivalent problem that is suitable for a solver that handles some generic class of problems, such as linear programs (LPs), quadratic programs (QPs), second-order cone programs (SOCPs), semidefinite programs (SDPs), and others such as exponential cone programs [7]. We focus on solvers that are suitable for embedded applications, i.e., are single-threaded, can be statically compiled, and do not make system calls: OSQP [14] handles QPs, SCS [15] and ECOS [16] handle cone programs that include SOCPs. After the canonicalized problem is solved, a solution of the original problem is retrieved from a solution of the canonicalized problem.

It is useful to think of the whole process as a function that maps  $\theta$ , the parameter that specifies the problem instance, into  $x^*(\theta)$ , an optimal value of the variable. With a DSL, this process consists of three steps. First the original problem description is canonicalized to a problem in some standard (or canonical) form; then the canonicalized problem is solved using a solver; and finally, a solution of the original problem is retrieved from a solution of the canonicalized problem.

Most of these DSLs are organized as *parser-solvers*, which carry out the canonicalization each time the problem is solved (with different parameter values). This simple setting is illustrated in Figure 1(a). We are interested in applications where we solve many instances of the problem, possibly in an embedded application with hard real-time constraints. For such applications, a *code generator* makes more sense. A code generator takes as input a description of a problem family, and generates specialized solver source code for that specific



(a) Parser-solver calculating solution  $x^*$  for problem instance with parameter  $\theta$ .

(b) Source code generation for problem family, followed by compilation to custom solver. The compiled solver computes a solution  $x^*$  to the problem instance with parameter  $\theta$ .

Fig. 1. Comparison of convex optimization problem parsing and solving approaches.

family. That source code is then compiled, and we have an efficient solver for the specific family.

This workflow is illustrated in Figure 1(b). The compiled solver has a number of advantages over parser-solvers. First, the compiled solver can be deployed in embedded systems, fulfilling rules for safety-critical code [17]. Second, by caching canonicalization and exploiting the problem structure, the compiled solver is faster.

A well known code generator is CVXGEN [18]. It handles problems that can be transformed to QPs and includes a custom interior-point solver. CVXGEN is used in many applications, including autonomous driving, dynamic energy management, and real-time trading in finance. SpaceX uses CVXGEN to generate flight code for high-speed onboard convex optimization for precision landing of space rockets [19].

CVXGEN was designed for use in real-time control systems, where the problems solved are not too big, either in terms of the number of variables or number of parameters. CVXGEN aggressively unrolls for-loops in its generated source code files, which means that every iteration of a for-loop is written as a separate instruction. This increases the solving speed, but can also result in large compiled code size. Due to the flat and explicit code generated, CVXGEN only handles problems with up to around a few thousand parameters. (More accurately, CVXGEN is limited to 4000 nonzero entries in the linear system of equations solved in each iteration.)

### A. Contribution

In this letter, we introduce the code generation tool CVXPYgen, which produces custom C code to solve a parameterized family of convex optimization problems. The design decisions for CVXPYgen are somewhat different from those made for CVXGEN. First, CVXPYgen is built on top of the DSL CVXPY, whereas CVXGEN is entirely self-contained. This means that prototypes can be developed, prototyped, and simulated in Python using CVXPY. Second, CVXPYgen interfaces with multiple solvers, currently OSQP, SCS, and ECOS. This means that CVXPYgen supports problems more

general than those that can be transformed to QPs. As far as we know, CVXPYgen is the first generic code generator for convex optimization that supports SOCPs. When using OSQP or SCS (both based on first-order methods), the generated solvers support warm-starting, which can bring more speed in some applications [4]. Third, CVXPYgen does not aggressively unroll loops in the generated code, which allows it to support high-dimensional parameters. In addition, matrix parameters can have any user-defined sparsity pattern. CVXPYgen uses partial update canonicalization, in which only the parameters changed are processed when solving a new problem instance. Fourth, CVXPYgen and its generated solvers are fully open-source, whereas CVXGEN is proprietary.

CVXPYgen (and more generally, code generation) is useful for two families of practical applications. The first is solving convex optimization problems in real-time settings on embedded devices, as is done in control systems, real-time resource allocators, and other applications. The second is in solving a large number of instances of a problem family, possibly on general-purpose computers. One example is back-testing in finance, where a trading policy based on convex optimization is simulated on historical or simulated data over many periods. Typical back-tests involve solving thousands or more instances of a problem family. In these applications, there is no hard real-time constraint; the goal is simply to speed up solving by avoiding repeatedly canonicalizing the problem.

### B. Prior Work

Several other code generators for optimization have been developed in addition to CVXGEN. FORCESPRO [20] and FORCES NLP [21] are proprietary code generators for multi-stage control problems. They handle problems that can be transformed to multi-stage quadratically constrained quadratic programs and nonlinear programs, respectively. The modular code generator acados [22] is open-source and specific to nonlinear optimal control problems. We focus on general convex problems up to SOCPs, as done by the open-source code generators QCML [23] and CVXPY-CODEGEN [24], which interface with ECOS. These early prototypes were developed before CVXPY included support for parameters and are no longer actively supported or maintained.

### C. Outline

The remainder of this letter is structured as follows. In Section II we describe, at a high level, how CVXPYgen works, and in Section III, we illustrate how it is used with a simple example. In Section IV and Section V we compare CVXPYgen to CVXGEN (for embedded use) and CVXPY (for general purpose use), respectively. We conclude this letter in Section VI.

## II. CVXPYGEN

CVXPYgen is based on the open-source Python-embedded DSL CVXPY. CVXPY handles many types of conic programs and certain types of nonconvex problems, whereas we focus on LPs, QPs, and SOCPs for code generation. CVXPY provides modeling instructions that follow the mathematical description for convex optimization problems. It ensures that the modeled

problems are convex, using disciplined convex programming (DCP). DCP is the process of constructing convex functions by assembling given base functions in mathematical expressions using a simple set of rules [25]. DCP ensures that the resulting problem is convex, and also, readily canonicalized to a standard form.

In DCP, parameters are treated as constants, optionally with specified sign, and there are no restrictions about how these constants appear in the expressions defining the problem family. The recently developed concept of disciplined parameterized programming (DPP) puts additional restrictions on how parameters can enter a problem description. If a problem family description is DPP-compliant, then canonicalization and retrieval can be represented as *affine* mappings [26]. Thus DPP-compliant problems are reducible to ASA-form, which stands for *Affine-Solve-Affine* [26]. This is the key property we exploit in CVXPYgen. More about the DCP and DPP rules can be found in the aforementioned papers, or at <https://www.cvxpy.org>.

After CVXPY has reduced the DPP-compliant problem to ASA-form, CVXPYgen extracts a sparse matrix  $C$  that canonicalizes the user-defined parameters  $\theta$  to the parameters  $\tilde{\theta}$  appearing in the standard form solver:

$$\tilde{\theta} = C \begin{bmatrix} \theta \\ 1 \end{bmatrix}.$$

CVXPYgen analyzes  $C$  to determine the user-defined parameters (i.e., components of  $\theta$ ) that every standardized form parameter depends on. This information is used when generating the custom solver, where only slices of the above mapping are computed if not all user-defined parameters are updated between solves. In addition, it is very useful to know the set of updated canonical parameters when using the OSQP solver or SCS, as detailed below.

In a similar way the retrieval of the solution  $x^*$  for the original problem from a solution  $\tilde{x}^*$  of the canonicalized problem is an affine mapping,

$$x^* = R \begin{bmatrix} \tilde{x}^* \\ 1 \end{bmatrix},$$

where  $R$  is a sparse matrix. Typically  $R$  is a selector matrix, with only one nonzero entry in each row, equal to one, in which case this step can be handled via simple pointers in C.

CVXPYgen generates allocation-, library-, and division-free C code for the canonicalization and retrieval steps, which in essence are nothing more than sparse matrix-vector multiplication, with some logic that exploits pointers or partial updates. Sparse matrices are stored in compressed sparse column format [27] and dense matrices are stored as vectors via column-major flattening.

Any solver can be used to solve the canonicalized problem, which provides the final link:

$$\tilde{x}^* = \mathcal{S}(\tilde{\theta}),$$

where  $\mathcal{S}$  denotes the mapping from the canonicalized parameters to a solution of the canonicalized problem. (We assume here that the problem instance is feasible, and that when there are multiple solutions, we simply pick one.) If available, CVXPYgen uses the canonical solver's code generation

method to produce C code for canonical solving. As of now, only OSQP provides this functionality [28]. Otherwise, the solver's C code is simply copied, possibly modified for use in embedded applications.

OSQP and SCS provide a set of C functions for updating their parameters. This way, when only canonical vector parameters are updated, the factorization of the linear system involved in the OSQP or SCS algorithms can be cached and re-used, which can lead to substantial speed up and division-free code. In the same way as only the parts of  $\tilde{\theta}$  are re-canonicalized that depend on the updated parts of  $\theta$ , only the OSQP or SCS update functions associated with these parameters are called before the canonicalized problem is solved.

The code and the full documentation for CVXPYgen with its generated solvers are available at <https://pypi.org/project/cvxpygen>.

### III. SIMPLE EXAMPLE

We consider the nonnegative least squares problem

$$\begin{aligned} & \text{minimize} \quad \|Gx - h\|_2^2 \\ & \text{subject to} \quad x \geq 0, \end{aligned} \quad (2)$$

where  $x \in \mathbf{R}^n$  is the variable and  $G \in \mathbf{R}^{m \times n}$ ,  $h \in \mathbf{R}^m$  are parameters, so  $\theta = (G, h)$ . We will canonicalize this to the standard form accepted by OSQP,

$$\begin{aligned} & \text{minimize} \quad \frac{1}{2} \tilde{x}^T P \tilde{x} + q^T \tilde{x} \\ & \text{subject to} \quad l \leq A \tilde{x} \leq u, \end{aligned} \quad (3)$$

where  $\tilde{x} \in \mathbf{R}^{\tilde{n}}$  is the canonical variable and all other symbols are canonical parameters, i.e.,  $\tilde{\theta} = (P, q, A, l, u)$ . (In this form, entries of  $l$  can be  $-\infty$ , and entries of  $u$  can be  $+\infty$ .)

The naïve canonicalization of (2) to (3) takes  $\tilde{x} = x$  and

$$P = 2G^T G, \quad q = 2G^T h, \quad A = I, \quad l = 0, \quad u = \infty.$$

In this canonicalization,  $\tilde{\theta}$  is not an affine function of  $\theta$ , since some entries of  $\tilde{\theta}$  are products of entries of  $\theta$ .

The canonicalization that uses DPP first expresses problem (2) as

$$\begin{aligned} & \text{minimize} \quad \|\tilde{x}_2\|_2^2 \\ & \text{subject to} \quad \tilde{x}_2 = G\tilde{x}_1 - h, \quad \tilde{x}_1 \geq 0, \end{aligned}$$

with variable  $\tilde{x} = (\tilde{x}_1, \tilde{x}_2)$ , where  $\tilde{x}_1 = x$  and  $\tilde{x}_2 \in \mathbf{R}^m$ . We can express this as (3) with parameters

$$\begin{aligned} P &= \begin{bmatrix} 0 & 0 \\ 0 & 2I \end{bmatrix}, \quad q = 0, \quad A = \begin{bmatrix} G & -I \\ I & 0 \end{bmatrix}, \\ l &= (h, 0), \quad u = (h, \infty), \end{aligned}$$

where the second part of  $u$  has  $\infty$  in every entry, i.e., there is no upper bound on the second part of  $A\tilde{x}$ . In this canonicalization,  $\tilde{\theta}$  is indeed an affine function of  $\theta$ . The retrieval map has the simple (linear) form  $x^* = [I \ 0]\tilde{x}^*$ .

We generate code for this problem as shown in Figure 2. The problem is modeled with CVXPY in lines 5–9. The actual code generation is done in line 11. The variable and parameters are named in lines 5–7 via their name attributes. These names are used for C variable and function naming.

```

1 import cvxpy as cp
2 from cvxpygen import cpg
3
4 # model problem
5 x=cp.Variable(n, name='x')
6 G=cp.Parameter((m,n), name='G')
7 h=cp.Parameter(m, name='h')
8 p=cp.Problem(cp.Minimize(cp.sum_squares(G*x-h)),
9             [x>=0])
10 # generate code
11 cpg.generate_code(p)

```

Fig. 2. Code generation for example (2). We assume that the dimensions  $m$  and  $n$  have been previously defined.

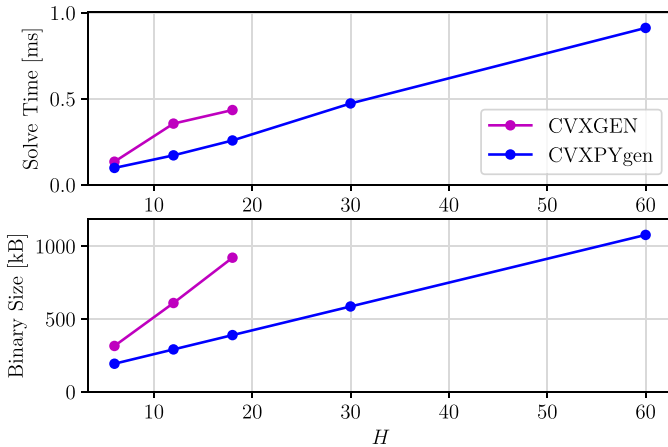


Fig. 3. Comparison of solve times (top) and binary sizes (bottom) with CVXGEN (magenta) and CVXPYgen (blue) used for code generation.

#### IV. COMPARISON TO CVXGEN

We compare CVXPYgen to CVXGEN for a model predictive control (MPC) problem family, described in Appendix A. In particular, we compare solve times of the C interface and executable sizes. The MPC problems are parameterized by their horizon length  $H \in \{6, 12, 18, 30, 60\}$ ; the number of variables is around  $10H$ . Figure 3 shows the resulting solve times averaged over 100 simulation steps, without warm-starting. CVXGEN is not able to generate code for  $H > 18$ , where the aforementioned limit of a few thousand parameters is exceeded. More precisely, there would be more than 4000 nonzero entries in the linear system of equations that is solved in every iteration. This limit is due to the very explicit coding style of CVXGEN. In contrast, the coding style of CVXPYgen is less explicit, e.g., making use of for-loops, thus longer horizons can be handled. CVXPYgen outperforms CVXGEN for all problem sizes. The automatically chosen OSQP solver is based on a very fast first-order method, whereas CVXGEN uses an interior-point method. The canonicalization and retrieval times with CVXPYgen are about two orders of magnitude smaller than the canonical solve times. Together, the overall solve times are smaller than with CVXGEN. In fact, enabling warm-starting with CVXPYgen (not possible with CVXGEN's interior-point method) can yield even more speed-up. The bottom of Figure 3 presents the example executable sizes for CVXGEN and CVXPYgen, respectively. For all values of  $H$ , the executables corresponding

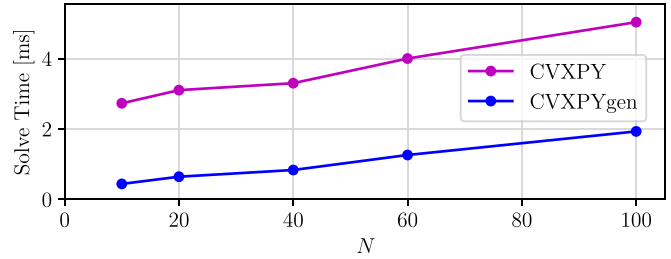


Fig. 4. Comparison of solve times with CVXPY (magenta) and the CVXPY interface of CVXPYgen (blue).

to CVXPYgen are considerably smaller. The main reason is the very explicit coding style of CVXGEN.

The execution times cited above are on a MacBook Pro 2.3GHz Intel i5. We have also used these generated solvers to control the position of a custom-built 14-by-14 cm quadcopter. The generated code was compiled in a robot operating system (ROS) node, and run on the drone's Intel Atom x5-Z8350 processor, at 30 Hz. We provide a video of the quadcopter following a circle trajectory at <https://bit.ly/3OXzOnq>.

#### V. COMPARISON TO CVXPY

Here we compare CVXPYgen to CVXPY, for a (financial) portfolio optimization problem family, described in Appendix B. This family of problems is parameterized by the number of assets in the portfolio  $N \in \{10, 20, 40, 60, 100\}$ . The number of variables in these problems is around  $2N$ . Figure 4 gives the results for 500 solves, a two year back-test using historical data. We see that the average solver speed is about 6 times faster with CVXPYgen, for  $N = 10$ , with the ratio dropping to 2.5 for  $N = 100$ . The execution times are measured on the same MacBook described in the previous section.

An interesting metric is the break-even point, which is the number of instances that need to be solved before CVXPYgen is faster than CVXPY, when we include the code generation and compilation time. This number is around 5000, and not too dependent on  $N$ . A typical back-test might involve daily trading, with around 250 trading days in each year, over 4 years, with hundreds of different hyper-parameter values, which gives on the order of 100,000 solves, well above this break-even point.

#### VI. CONCLUSION

We have described CVXPYgen, a tool for generating custom C code that solves instances of a family of convex optimization problems specified within CVXPY. This gives a seamless path from prototyping an application using Python and CVXPY, to a final embedded implementation in C. In addition to CVXPYgen supporting a wider variety of problems (such as SOCPs) than the state-of-the-art code generator CVXGEN, numerical experiments show that it outperforms CVXGEN in terms of allowable problem size, compiled code size, and solve times. For applications running on general purpose machines, we obtain a significant speedup over CVXPY when many problem instances are to be solved. In the future, it might be of interest to optimize the canonical problem

representations for solve times or binary sizes, while retaining ASA-representability.

## APPENDIX A MPC EXAMPLE

We use MPC to track the position and velocity of a quadcopter with mass  $m$ , experiencing gravitational acceleration  $g$ . We model the quadcopter as a point mass with position error  $p_k \in \mathbf{R}^3$  and velocity error  $v_k \in \mathbf{R}^3$ , where  $k$  denotes the time step or period. We concatenate the position and velocity error to the state  $z_k = (p_k, v_k) \in \mathbf{R}^6$ , which we regulate to 0. The input is the force vector  $u_k \in \mathbf{R}^3$  without gravity compensation. The dynamics are

$$z_{k+1} = Az_k + Bu_k,$$

where  $A \in \mathbf{R}^{6 \times 6}$  and  $B \in \mathbf{R}^{6 \times 3}$ .

We limit the tilt angle of the quadcopter. Since the quadcopter's attitude is tied to the pointing direction of  $u$  plus gravity compensation, we impose a (polyhedral) tilt angle constraint as

$$c_j^T (u_k)_{0:1} \leq \gamma((u_k)_2 + mg), \quad j = 0, \dots, N^{\text{hs}} - 1,$$

where  $(\cdot)_{0:1}$  and  $(\cdot)_2$  denote the horizontal and vertical part of a vector in  $\mathbf{R}^3$  space, respectively, and  $\gamma > 0$ . We use  $N^{\text{hs}}$  halfspaces parameterized through  $c_j$ . Compared to a spherical (natural) tilt angle constraint, when added to the MPC constraints, this formulation renders the problem QP representable. The lower and upper thrust limits of the propellers are represented as  $u_{\text{vmin}} \leq (u_k)_2 \leq u_{\text{vmax}}$  with  $u_{\text{vmin}} < 0$  and  $u_{\text{vmax}} > 0$ .

Up to horizon  $H$ , we penalize state errors and control effort via the traditional quadratic cost

$$z_H^T Q_T z_H + \sum_{k=0}^{H-1} (z_k^T Q z_k + u_k^T R u_k),$$

with diagonal positive definite matrices  $Q$  and  $R$ , and positive definite  $Q_T$ , which is the solution to the discrete-time algebraic Riccati equation (as a function of  $A$ ,  $B$ ,  $Q$ , and  $R$ ). In addition, at every stage, we discourage rapid changes of the input (that the low-level attitude control system cannot follow) with the additional cost term

$$\sum_{k=0}^{H-1} (u_{k+1} - u_k)^T T (u_{k+1} - u_k),$$

where  $T$  is diagonal positive definite. Combining all the above constraints and cost terms, we arrive at the MPC problem

$$\begin{aligned} \text{minimize} \quad & z_H^T Q_T z_H + \sum_{k=0}^{H-1} (z_k^T Q z_k + u_k^T R u_k + \\ & (u_{k+1} - u_k)^T T (u_{k+1} - u_k)) \\ \text{subject to} \quad & z_0 = z_{\text{meas}}, \quad u_0 = u_{\text{prev}} \\ & z_{k+1} = Az_k + Bu_k, \quad k = 0, \dots, H-1 \\ & u_{\text{vmin}} \leq (u_k)_2 \leq u_{\text{vmax}}, \quad k = 1, \dots, H-1 \\ & c_j^T (u_k)_{0:1} \leq \gamma((u_k)_2 + mg), \\ & j = 1, \dots, N^{\text{hs}} - 1, \quad k = 1, \dots, H-1, \end{aligned}$$

where the states  $z_k$  and inputs  $u_k$  are optimization variables. The current state measurement is  $z_{\text{meas}}$  and the solution for the input at the first stage from the previous solve is  $u_{\text{prev}}$ . In practice, these two would most certainly be the only parameters of the problem. However, for demonstration purposes, we declare all other symbols (except for variables, all  $c_j$ ,  $N^{\text{hs}}$ , and  $H$ ) as parameters. This problem formulation is not DPP-compliant, e.g., because of the multiplication of parameters  $\gamma$ ,  $m$ , and  $g$ .

Before rewriting the problem in DPP-compliant form, we define the following convenience notation for matrix slicing. We use the zero-based counting scheme.  $M_{r:t,c:d}$  is the slice of some matrix  $M$  from its  $r$ th to its  $t$ th row (included) and from its  $c$ th to its  $d$ th column (included). We slice full columns by omitting the row indices, i.e.,  $M_c$  is the  $c$ th column of  $M$ . Finally, we write the DPP-compliant problem as

$$\begin{aligned} \text{minimize} \quad & \|Q_T^{1/2} Z_H\|_2^2 + \|Q^{1/2} Z_{0:H-1}\|_F^2 \\ & + \|R^{1/2} U_{0:H-1}\|_F^2 + \|T^{1/2} (U_{1:H} - U_{0:H-1})\|_F^2 \\ \text{subject to} \quad & Z_0 = z_{\text{meas}}, \quad U_0 = u_{\text{prev}} \\ & Z_{1:H} = AZ_{0:H-1} + BU_{0:H-1} \\ & u_{\text{vmin}} \leq U_{2,1:H-1} \leq u_{\text{vmax}} \\ & c_j^T U_{0:1,1:H-1} \leq \gamma U_{2,1:H-1} + d, \\ & j = 0, \dots, N^{\text{hs}} - 1, \end{aligned}$$

where  $Z \in \mathbf{R}^{6 \times (H+1)}$  and  $U \in \mathbf{R}^{3 \times (H+1)}$  are the variables and contain the state and input vectors, respectively, for increasing stage count in their columns. The vector  $d \in \mathbf{R}^{H-1}$  contains  $\gamma mg$  in all its entries. The problem is parameterized by

$$Q_T^{1/2}, Q^{1/2}, R^{1/2}, T^{1/2}, A, B, \\ \gamma, d, u_{\text{vmin}}, u_{\text{vmax}}, z_{\text{meas}}, u_{\text{prev}}.$$

In the expressions above,  $M^{1/2}$  denotes any squareroot of the positive definite matrix  $M$ , e.g., the transposed Cholesky factor,  $\|\cdot\|_F$  denotes the Frobenius norm, and the inequalities are elementwise.

## APPENDIX B PORTFOLIO OPTIMIZATION EXAMPLE

We search for a portfolio of holdings in  $N$  assets and a cash balance. The corresponding weights are  $w \in \mathbf{R}^{N+1}$ , where the last entry represents the cash balance, and  $\mathbf{1}^T w = 1$ . We impose a leverage limit  $\|w\|_1 \leq L$ , where  $L \geq 1$  is a parameter.

The return  $r \in \mathbf{R}^{N+1}$  has mean (or forecast)  $\alpha \in \mathbf{R}^{N+1}$ , so the expected portfolio return is  $\alpha^T w$ . The risk or variance of the portfolio return is  $w^T \Sigma w$ , where  $\Sigma$  is the positive definite covariance matrix of the asset returns.

We consider two additional objective terms. One is a trading or transaction cost  $(\kappa^{\text{tc}})^T |w - w^{\text{prev}}|$ , where the absolute value is elementwise,  $w^{\text{prev}}$  is the previous period weights, and  $\kappa^{\text{tc}} \geq 0$  is a parameter. The other is a short-selling cost  $(\kappa^{\text{sh}})^T (w)_-$ , where  $(w)_- = \max\{-w, 0\}$  (elementwise), and  $\kappa^{\text{sh}} \geq 0$  is a parameter.

The overall objective function is

$$\alpha^T w - \gamma^{\text{risk}} w^T \Sigma w - \gamma^{\text{tc}} (\kappa^{\text{tc}})^T |w - w^{\text{prev}}| - \gamma^{\text{sh}} (\kappa^{\text{sh}})^T (w)_-,$$

where  $\gamma^{\text{risk}}$ ,  $\gamma^{\text{tc}}$ , and  $\gamma^{\text{sh}}$  are positive parameters that scale the risk, transaction cost, and shorting cost, respectively.

Our final optimization problem is

$$\begin{aligned} & \text{maximize} \quad \alpha^T w - \gamma^{\text{risk}} w^T \Sigma w \\ & \quad - \gamma^{\text{tc}} (\kappa^{\text{tc}})^T |w - w^{\text{prev}}| - \gamma^{\text{sh}} (\kappa^{\text{sh}})^T (w)_- \\ & \text{subject to} \quad \mathbf{1}^T w = 1, \quad \|w\|_1 \leq L, \end{aligned}$$

where  $w \in \mathbf{R}^{N+1}$  is the variable, and all other symbols are parameters.

The covariance matrix  $\Sigma$  takes the standard factor model form,

$$\Sigma = FF^T + D,$$

where  $F \in \mathbf{R}^{(N+1) \times K}$  and  $D$  is positive definite diagonal. The number of factors in this risk model is  $K$ , which is usually much less than  $N$ .

This problem formulation is not DPP-compliant, but we can rewrite it in DPP-compliant form by eliminating quadratic forms and collecting products of parameters, as in

$$\begin{aligned} & \text{maximize} \quad \left( \frac{\alpha}{\gamma^{\text{risk}}} \right)^T w - \|F^T w\|_2^2 - \|D^{1/2} w\|_2^2 \\ & \quad - \left( \frac{\gamma^{\text{tc}}}{\gamma^{\text{risk}}} \kappa^{\text{tc}} \right)^T |\Delta w| - \left( \frac{\gamma^{\text{sh}}}{\gamma^{\text{risk}}} \kappa^{\text{sh}} \right)^T (w)_- \\ & \text{subject to} \quad \mathbf{1}^T w = 1, \quad \|w\|_1 \leq L \\ & \quad \Delta w = w - w^{\text{prev}}, \end{aligned}$$

where the portfolio weight vector  $w \in \mathbf{R}^{N+1}$  and the weight change vector  $\Delta w \in \mathbf{R}^{N+1}$  are the variables. The problem is parameterized by

$$\frac{\alpha}{\gamma^{\text{risk}}}, \quad F, \quad D^{1/2}, \quad \frac{\gamma^{\text{tc}}}{\gamma^{\text{risk}}} \kappa^{\text{tc}}, \quad \frac{\gamma^{\text{sh}}}{\gamma^{\text{risk}}} \kappa^{\text{sh}}, \quad L, \quad w^{\text{prev}}.$$

We consider  $N$  stock assets, chosen randomly from the S&P 500, with historical return data from 2017–2019. For each value of  $N$ , we set  $K = \max(N/10, 5)$ .

## ACKNOWLEDGMENT

The authors would like to express their gratitude to John Lygeros for enabling this collaboration. Moreover, the authors would like to thank JunEn Low and Mac Schwager for providing the quadcopter testing environment at the Stanford Flight Room. M. Schaller received the scholarship of the “German Academic Scholarship Foundation” for his studies at ETH, where he conducted the research presented in this paper.

## REFERENCES

- [1] J. Mattingley and S. Boyd, “Real-time convex optimization in signal processing,” *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 50–61, May 2010.
- [2] M. Zibulevsky and M. Elad, “L1-L2 optimization in signal and image processing,” *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 76–88, May 2010.
- [3] Y. Wang and S. Boyd, “Fast model predictive control using online optimization,” *IEEE Trans. Control Syst. Technol.*, vol. 18, no. 2, pp. 267–278, Mar. 2010.
- [4] C. E. García, D. M. Prett, and M. Morari, “Model predictive control: Theory and practice—A survey,” *Automatica*, vol. 25, no. 3, pp. 335–348, 1989.
- [5] S. Boyd *et al.*, “Multi-period trading via convex optimization,” *Foundations Trends Optimization*, vol. 3. Hanover, MA, USA: Now Publ., 2017, pp. 1–76.
- [6] H. Markowitz, “Portfolio selection,” *J. Finan.*, vol. 7, no. 1, pp. 77–91, 1952.
- [7] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [8] J. Löfberg, “YALMIP : A toolbox for modeling and optimization in MATLAB,” in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, 2004, pp. 284–289.
- [9] M. Grant and S. Boyd, *CVX: MATLAB Software for Disciplined Convex Programming, Version 2.1*, CVX Res., Austin, TX, USA, 2014.
- [10] S. Diamond and S. Boyd, “CVXPY: A python-embedded modeling language for convex optimization,” *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 2909–2913, 2016.
- [11] M. Udell, K. Mohan, D. Zeng, J. Hong, S. Diamond, and S. Boyd, “Convex optimization in Julia,” in *Proc. SC Workshop High Perform. Tech. Comput. Dyn. Lang.*, 2014, pp. 18–28.
- [12] I. Dunning, J. Huchette, and M. Lubin, “JuMP: A modeling language for mathematical optimization,” *SIAM Rev.*, vol. 59, no. 2, pp. 295–320, 2017.
- [13] A. Fu, B. Narasimhan, and S. Boyd, “CVXR: An R package for disciplined convex optimization,” *J. Stat. Softw.*, vol. 94, no. 14, pp. 1–34, 2020.
- [14] B. Stellato, G. Banjac, P. Goulart, A. Bemporad, and S. Boyd, “OSQP: An operator splitting solver for quadratic programs,” *Math. Program. Comput.*, vol. 12, no. 4, pp. 637–672, 2020.
- [15] B. O’Donoghue, E. Chu, N. Parikh, and S. Boyd, “Conic optimization via operator splitting and homogeneous self-dual embedding,” *J. Optim. Theory Appl.*, vol. 169, no. 3, pp. 1042–1068, 2016.
- [16] A. Domahidi, E. Chu, and S. Boyd, “ECOS: An SOCP solver for embedded systems,” in *Proc. IEEE Eur. Control Conf. (ECC)*, 2013, pp. 3071–3076.
- [17] G. Holzmann, “The power of 10: Rules for developing safety-critical code,” *Computer*, vol. 39, no. 6, pp. 95–99, 2006.
- [18] J. Mattingley and S. Boyd, “CVXGEN: A code generator for embedded convex optimization,” *Optim. Eng.*, vol. 13, no. 1, pp. 1–27, 2012.
- [19] L. Blackmore, “Autonomous precision landing of space rockets,” in *Proc. Front. Eng. Rep. Leading-Edge Eng. Symp.*, Washington, DC, USA, 2016, pp. 15–20.
- [20] A. Domahidi and J. Jerez. “FORCES Professional.” Embotech AG. 2019. [Online]. Available: <https://embotech.com/FORCES-Pro>
- [21] A. Zanelli, A. Domahidi, J. Jerez, and M. Morari, “FORCES NLP: An efficient implementation of interior-point methods for multistage nonlinear nonconvex programs,” *Int. J. Control*, vol. 93, no. 1, pp. 13–29, 2020.
- [22] R. Verschuere *et al.*, “acados—A modular open-source framework for fast embedded optimal control,” *Math. Program. Comput.*, vol. 14, pp. 147–183, Oct. 2021.
- [23] E. Chu and S. Boyd. “QCML: Quadratic Cone Modeling Language.” 2017. [Online]. Available: <https://github.com/cvxgrp/qcml>
- [24] N. Moehle, J. Mattingley, and S. Boyd. “Embedded Convex Optimization with CVXPY.” 2017. [Online]. Available: [https://github.com/moehle/cvxpy\\_codegen](https://github.com/moehle/cvxpy_codegen)
- [25] M. Grant, “Disciplined convex programming,” Dept. Electr. Eng., Stanford Univ., Stanford, CA, USA, 2004.
- [26] A. Agrawal, B. Amos, S. Barratt, S. Boyd, S. Diamond, and Z. Kolter, “Differentiable convex optimization layers,” in *Advances in Neural Information Processing System (NeurIPS)*. Red Hook, NY, USA: Curran, 2019.
- [27] A. Buluç, J. T. Fineman, M. Frigo, J. R. Gilbert, and C. E. Leiserson, “Parallel sparse matrix-vector and matrix-transpose-vector multiplication using compressed sparse blocks,” in *Proc. Symp. Parallelism Algorithms Archit. (SPAA)*, 2009, pp. 233–244.
- [28] G. Banjac, B. Stellato, N. Moehle, P. Goulart, A. Bemporad, and S. Boyd, “Embedded code generation using the OSQP solver,” in *Proc. IEEE Conf. Decis. Control (CDC)*, 2017, pp. 1906–1911.