copters: Algorithms Reference

Avinash Madavan

12. February 2026

This document is a description of the algorithms applied in copters. It is not intended to be a comprehensive treatment of optimization theory, but rather a dive into the implementation and key concepts used in the software. We will attempt to provide references to more complete treatments of the topics discussed here, but do not guarantee their completeness.

Contents

1 Introduction

Copters is a library for solving optimization problems, both linear and nonlinear. In this section, we will introduce the basic concepts of optimization theory that are relevant to the algorithms implemented in copters. We will discuss the formulation of optimization problems, as well as key concepts that will be used to describe the algorithms.

We begin with a general formulation of optimization problems, and then specialize to linear programming (LP) and nonlinear programming (NLP). This general form seeks to optimize (minimize or maximize) an objective function subject to constraints on the decision variables. The general (non-linear) optimization problem can be stated as:

$$
\begin{aligned}
\underset{\boldsymbol{x}}{\text{minimize}} \quad & f(\boldsymbol{x}), \\
\text{subject to,} \quad & \boldsymbol{g}(\boldsymbol{x}) = 0, \\
& \boldsymbol{h}(\boldsymbol{x}) \leq 0, \\
& \boldsymbol{x} \in \mathcal{X},
\end{aligned} \tag{1.1}
$$

where $\boldsymbol{f}$ is the objective function, $\boldsymbol{g}$ are the equality constraints, $\boldsymbol{h}$ are the inequality constraints, and $\boldsymbol{X}$ represents the feasible set for the decision variables.

TODO: Add discussion of duality.

We associate dual variables $\boldsymbol{y}$, $\boldsymbol{z}$, and $\boldsymbol{w}$ with the equality constraints, inequality constraints, and set membership constraints, respectively. The associated Lagrangian is given by:

$$
\mathcal{L}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \geq 0, \boldsymbol{w} \in W) = f(\boldsymbol{x}) - \boldsymbol{y}^T \boldsymbol{g}(\boldsymbol{x}) - \boldsymbol{z}^T \boldsymbol{h}(\boldsymbol{x}) - I_{\{\mathcal{W}\}}(\boldsymbol{x}). \tag{1.2}
$$

The problem Equation (1.1) can be equivalently written as:

$$
\min_{\boldsymbol{x}} \max_{(\boldsymbol{y}, \boldsymbol{z} \geq 0, \boldsymbol{w} \in W)} \mathcal{L}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{w}) \tag{1.3}
$$

where the maximization is over the dual variables.It can then be shown that we can construct the lower bound:

$$
\min_{\boldsymbol{x}} \max_{(\boldsymbol{y}, \boldsymbol{z} \geq 0, \boldsymbol{w} \in \mathcal{W})} \mathcal{L}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{w}) \geq \max_{(\boldsymbol{y}, \boldsymbol{z} \geq 0, \boldsymbol{w} \in \mathcal{W})} \min_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{w}). \tag{1.4}
$$

The lower bound is itself an optimization problem and is referred to as the dual problem. The weak duality theorem states that the optimal value of the primal problem is always greater than or equal to the optimal value of the dual problem. Under certain conditions, known as strong duality conditions, the two optimal values are equal.

The dual problem can be equivalently written as:

$$
\begin{aligned}
\underset{(\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{w})}{\text{maximize}} \quad & \min_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{w}), \\
\text{subject to,} \quad & \boldsymbol{z} \geq 0, \boldsymbol{w} \in \mathcal{W}.
\end{aligned} \tag{1.5}
$$

The dual problem can be used to derive necessary and sufficient conditions for optimality of the primal problem, known as the Karush-Kuhn-Tucker (KKT) conditions, which require satisfaction of the primal feasibility, dual feasibility, complementary slackness, and stationarity conditions. The complementary slackness condition ensures that for each inequality constraint, either the constraint is active (i.e., holds with equality) or the corresponding dual variable is zero. The KKT conditions for optimality of the primal and dual problems are given by:

$$\begin{aligned}
&\boldsymbol{g}(\boldsymbol{x}) = 0, \\
&\boldsymbol{h}(\boldsymbol{x}) \leq 0, \\
&\nabla f(\boldsymbol{x}) - \nabla \boldsymbol{g}(\boldsymbol{x})^T \boldsymbol{y} - \nabla \boldsymbol{h}(\boldsymbol{x})^T \boldsymbol{z} - \nabla I_{\{\mathcal{W}\}}(\boldsymbol{x}) = 0, \\
&\boldsymbol{z}^T \boldsymbol{h}(\boldsymbol{x}) = 0, \\
&\boldsymbol{z} \geq 0, \boldsymbol{x} \in \mathcal{X},
\end{aligned} \tag{1.6}$$

where the first two equations represent primal feasibility, the third equation represents stationarity, the fourth equation represents complementary slackness, and the last two equations represent dual feasibility.

In the following subsections we will explore specific cases of optimization problems, namely linear programming (LP) and nonlinear programming (NLP), and derive their respective dual problems and KKT conditions. In these instances the above equations may simplify due to the specific structure of the objective functions and constraints.

1.1 Linear Programming (LP)

The LP problems is a special case of the general optimization problem where the objective function and constraints are all linear. A standard form of a linear programming problem can be stated as:

$$\begin{aligned}
\underset{\boldsymbol{x}}{\text{minimize}} \quad & \boldsymbol{c}^T \boldsymbol{x}, \\
\text{subject to,} \quad & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \quad && : \boldsymbol{y} \\
& \underline{\boldsymbol{x}} \leq \boldsymbol{x} \leq \overline{\boldsymbol{x}}. && : \underline{\boldsymbol{w}}, \overline{\boldsymbol{w}}
\end{aligned} \tag{1.1.1}$$

Here, we have associated dual variables, y, $\underline{w}$, and $\overline{w}$ with the linear constraint, lower bound, and upper bound, respectively. In this problem the objective function is now a linear function $f(\boldsymbol{x}) = \boldsymbol{c}^T \boldsymbol{x}$, the equality constraints are given by $\boldsymbol{g}(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}$, and the inequality constraint is ignored. It can be shown that inequality constraints can easily be reformulated as equality constraints. The set membership constraint restricts $\mathcal{X} = [\underline{\boldsymbol{x}}, \overline{\boldsymbol{x}}]$.

We can then construct the associated Lagrangian

$$\mathscr{L}(\boldsymbol{x}, \boldsymbol{y}, \underline{\boldsymbol{w}} \geq 0, \overline{\boldsymbol{w}} \leq 0) = \boldsymbol{c}^T \boldsymbol{x} - \boldsymbol{y}^T (\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}) - \underline{\boldsymbol{w}}^T (\boldsymbol{x} - \underline{\boldsymbol{x}}) - \overline{\boldsymbol{w}}^T (\boldsymbol{x} - \overline{\boldsymbol{x}}) \tag{1.1.2}$$

The dual problem can then be constructed as:

$$
\begin{aligned}
\underset{(\boldsymbol{y},\underline{\boldsymbol{w}}\geq 0,\overline{\boldsymbol{w}}\leq 0)}{\text{maximize}} \quad & \boldsymbol{b}^T\boldsymbol{y} + \underline{\boldsymbol{w}}^T\underline{\boldsymbol{x}} + \overline{\boldsymbol{w}}^T\overline{\boldsymbol{x}}, \\
\text{subject to,} \quad & A^T y + \underline{z} - \overline{z} = c, \\
& \underline{\boldsymbol{w}} \geq 0, \overline{\boldsymbol{w}} \leq 0.
\end{aligned} \tag{1.1.3}
$$

Necessary and sufficient conditions for optimality of the primal and dual problems are given by the KKT conditions:

$$
\begin{aligned}
& \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \\
& \boldsymbol{c} - \boldsymbol{A}^T\boldsymbol{y} - \underline{\boldsymbol{w}} + \overline{\boldsymbol{w}} = 0, \\
& \underline{\boldsymbol{w}}^T(\boldsymbol{x} - \underline{\boldsymbol{x}}) = 0, \overline{\boldsymbol{w}}^T(\boldsymbol{x} - \overline{\boldsymbol{x}}) = 0, \\
& \underline{\boldsymbol{x}} \leq \boldsymbol{x} \leq \overline{\boldsymbol{x}}, \underline{\boldsymbol{w}} \geq 0, \overline{\boldsymbol{w}} \leq 0,
\end{aligned} \tag{1.1.4}
$$

where the first two equations are the primal and dual feasibility conditions, the next three equations are the complementary slackness conditions, and the last two equations are the dual feasibility conditions.

1.2 Nonlinear Programming (NLP)

A general nonlinear programming (NLP) problem can be stated as:

$$
\begin{aligned}
\underset{x}{\text{minimize}} \quad & f(x), \\
\text{subject to,} \quad & g(x) = 0, : y, \\
& xin[l,u], \ : \underline{z}, \overline{z}
\end{aligned} \tag{1.2.1}
$$

where $f : R^n \to R$ is the objective function, $g : R^n \to R^m$ are the equality constraints, and $x \in [l, u]$ represents the bound constraints on the decision variables. The associated Langrangian is given by:

$$
\mathscr{L}(x, y, \underline{z} \geq 0, \overline{z} \leq 0) = f(x) - y^T g(x) - \underline{z}^T(x - l) - \overline{z}^T(x - u). \tag{1.2.2}
$$

The KKT conditions for optimality are given by:

$$
\begin{aligned}
& g(x) = 0, \\
& \nabla f(x) - \nabla g(x)^T y - \underline{z} + \overline{z} = 0, \\
& \underline{z}^T(x - l) = 0, \overline{z}^T(x - u) = 0, \\
& x \in [l, u], \underline{z} \geq 0, \overline{z} \leq 0,
\end{aligned} \tag{1.2.3}
$$

where the first equation represents primal feasibility, the second equation represents dual feasibility, the next two equations represent complementary slackness, and the last three equations represent the bound constraints on the primal and dual variables.

The KKT conditions Equation (1.2.3) are necessary for optimality under certain regularity conditions, such as the Mangasarian-Fromovitz constraint qualification (MFCQ). They are sufficient for optimality if the objective function f is convex and the constraint functions g are affine. They are sufficient for local optimality if f and g are twice continuously differentiable and the second-order sufficient conditions hold, given by the positive definiteness of the Lagrangian Hessian on the critical cone.

2 Linear Programming algorithms

2.1 Mehrotra Predictor-Corrector

The Mehrotra Predictor-Corrector is an example of a primal-dual interior point method for solving linear programs. While it does not guarantee polynomial-time convergence, in practive it typically demonstrates convergence in a small number of iterations.

The theory behind this algorithm follows from the general primal-dual interior point framework, which is described in a later section (TODO: link). For the purposes of this discussion, we restrict attention to linear programming and the application of Mehrotra's method to this problem class.

The KKT conditions Equation (1.1.4) for the linear program Equation (1.1.1) can be rearranged to give the following system of equations:

$$
\begin{aligned}
&\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} = 0,\\
&\boldsymbol{c} - \boldsymbol{A}^T\boldsymbol{y} - \underline{\boldsymbol{z}} - \overline{\boldsymbol{z}} = 0,\\
&(\boldsymbol{X} - \boldsymbol{L})\underline{\boldsymbol{Z}}\boldsymbol{e} = \sigma\mu\boldsymbol{e},\\
&(\boldsymbol{X} - \boldsymbol{U})\overline{\boldsymbol{Z}}\boldsymbol{e} = \sigma\mu\boldsymbol{e},
\end{aligned}
\tag{2.1.1}
$$

where $\boldsymbol{e}$ is the vector of all ones, and the matrices $\boldsymbol{X}$, $\boldsymbol{L}$, $\boldsymbol{U}$, $\underline{\boldsymbol{Z}}$, and $\overline{\boldsymbol{Z}}$ are diagonal matrices with the elements of the corresponding vectors on their diagonals. The parameter μ is used to ensure positivity of complimentary slackness conditions. It can be seen that this is equivalent to Equation (1.1.4) for $\mu = 0$.

The Mehrotra Predictor-Corrector algorithm aims to solve a linearization of the above system, while simultaneosly driving μ to zero. The linearization is given by:

$$
\begin{aligned}
&\boldsymbol{A}(\boldsymbol{x}+\Delta\boldsymbol{x})-\boldsymbol{b}=0,\\
&\boldsymbol{c}-\boldsymbol{A}^T(\boldsymbol{y}+\Delta\boldsymbol{y})-(\underline{\boldsymbol{z}}+\Delta\underline{\boldsymbol{z}})-(\overline{\boldsymbol{z}}+\Delta\overline{\boldsymbol{z}})=0,\\
&(\boldsymbol{X}-\boldsymbol{L})\underline{\boldsymbol{Z}}\boldsymbol{e}+(\boldsymbol{X}-\boldsymbol{L})\Delta\underline{\boldsymbol{z}}+\underline{\boldsymbol{Z}\Delta\boldsymbol{x}}=\sigma\mu\boldsymbol{e},\\
&(\boldsymbol{X}-\boldsymbol{U})\overline{\boldsymbol{Z}}\boldsymbol{e}+(\boldsymbol{X}-\boldsymbol{U})\Delta\overline{\boldsymbol{z}}+\overline{\boldsymbol{Z}\Delta\boldsymbol{x}}=\sigma\mu\boldsymbol{e}.
\end{aligned}
\tag{2.1.2}
$$

Collecting constant terms on the right-hand side, this yields the following system of equations in the search directions $\Delta\boldsymbol{x}$, $\Delta\boldsymbol{y}$, $\Delta\underline{\boldsymbol{z}}$, and $\Delta\overline{\boldsymbol{z}}$:

$$
\begin{aligned}
&\boldsymbol{A}\Delta\boldsymbol{x}=\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x},\\
&\boldsymbol{A}^T\Delta\boldsymbol{y}+\Delta\underline{\boldsymbol{z}}+\Delta\overline{\boldsymbol{z}}=\boldsymbol{c}-\boldsymbol{A}^T\boldsymbol{y}-\underline{\boldsymbol{z}}-\overline{\boldsymbol{z}},\\
&(\boldsymbol{X}-\boldsymbol{L})\Delta\underline{\boldsymbol{z}}+\underline{\boldsymbol{Z}}\Delta\boldsymbol{x}=\sigma\mu\boldsymbol{e}-(\boldsymbol{X}-\boldsymbol{L})\underline{\boldsymbol{Z}}\boldsymbol{e},\\
&(\boldsymbol{X}-\boldsymbol{U})\Delta\overline{\boldsymbol{z}}+\overline{\boldsymbol{Z}}\Delta\boldsymbol{x}=\sigma\mu\boldsymbol{e}-(\boldsymbol{X}-\boldsymbol{U})\overline{\boldsymbol{Z}}\boldsymbol{e}.
\end{aligned}
\tag{2.1.3}
$$

The linearization of the system is given by:

$$
\begin{aligned}
&\boldsymbol{A}\Delta\boldsymbol{x}=\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x},\\
&\boldsymbol{A}^T\Delta\boldsymbol{y}+\Delta\underline{\boldsymbol{z}}+\Delta\overline{\boldsymbol{z}}=\boldsymbol{c}-\boldsymbol{A}^T\boldsymbol{y}-\underline{\boldsymbol{z}}-\overline{\boldsymbol{z}},\\
&(\boldsymbol{X}-\boldsymbol{L})\Delta\underline{\boldsymbol{z}}+\underline{\boldsymbol{Z}}\Delta\boldsymbol{x}=\sigma\mu\boldsymbol{e}-(\boldsymbol{X}-\boldsymbol{L})\underline{\boldsymbol{Z}}\boldsymbol{e},\\
&(\boldsymbol{X}-\boldsymbol{U})\Delta\overline{\boldsymbol{z}}+\overline{\boldsymbol{Z}}\Delta\boldsymbol{x}=\sigma\mu\boldsymbol{e}-(\boldsymbol{X}-\boldsymbol{U})\overline{\boldsymbol{Z}}\boldsymbol{e}.
\end{aligned}
\tag{2.1.4}
$$

The algorithm proceeds by solving the above linear system for the search directions $\Delta\boldsymbol{x}$, $\Delta\boldsymbol{y}$, $\Delta\underline{\boldsymbol{z}}$, and $\Delta\overline{\boldsymbol{z}}$. The step sizes are then computed to ensure that the updated variables remain feasible with respect to the bound constraints. Finally, the variables are updated using the computed step sizes and search directions.

The key innovation of Mehrotra's method is the use of a predictor-corrector approach to improve convergence. In the predictor step, the algorithm solves the linear system with $\sigma = 0$ to obtain an initial estimate of the search directions. The value of μ is then updated based on the predicted step, and the center-corrector step is performed by solving the linear system again with the updated value of σ. (TODO: update with information for centering and corrector description)

Suppose the above system results in a search direction $\Delta\boldsymbol{x}_p$, $\Delta\boldsymbol{y}_p$, $\Delta\underline{\boldsymbol{z}}_p$, and $\Delta\overline{\boldsymbol{z}}_p$ in the predictor step. The algorithm must then ensure that the subsequent iteration remains feasible with respect to the bound constraints. We thus require step sizes α_p and α_d that ensure positivity. Taking the primal lower bound as an example, we require:

$$
\boldsymbol{x}+\alpha_p\Delta\boldsymbol{x}\geq\underline{\boldsymbol{x}}
\tag{2.1.5}
$$

In the case that $\Delta \boldsymbol{x}_{\{p,i\}} \geq 0$, we know that $\boldsymbol{x} \geq \underline{\boldsymbol{x}}$ and $\alpha_p = 0$, so the above condition is always satisfied. Thus, we must only consider the case where $\Delta \boldsymbol{x}_{\{p,i\}} < 0$. Rearranging the above condition gives:

$$\alpha_p \leq \frac{\underline{\boldsymbol{x}}_i - \boldsymbol{x}_i}{\Delta \boldsymbol{x}_i} \quad \forall i, \Delta \boldsymbol{x}_i < 0. \tag{2.1.6}$$

We ensure this condition by ensuring that α_p is less than or equal to the minimum of the right-hand side over all i where $\Delta \boldsymbol{x}_i < 0$. To provide some margin, we typically scale this value by a factor τ in $(0, 1)$. A similar process can be used to identify the dual step size α_d.

The system of equations in Equation (2.1.4) can be reduced to variables in $\Delta \boldsymbol{x}, \Delta \boldsymbol{y}$, by noting that $\Delta \underline{\boldsymbol{z}}$ and $\Delta \overline{\boldsymbol{z}}$ can be expressed in terms of $\Delta \boldsymbol{x}$. Specifically, we have:

$$\begin{aligned}
\Delta \underline{\boldsymbol{z}} &= (\boldsymbol{X} - \boldsymbol{L})^{-1}(\sigma\mu\boldsymbol{e} - (\boldsymbol{X} - \boldsymbol{L})\underline{\boldsymbol{Z}}\boldsymbol{e} - \underline{\boldsymbol{Z}}\Delta\boldsymbol{x}), \\
&= (\boldsymbol{X} - \boldsymbol{L})^{-1}\sigma\mu\boldsymbol{e} - \underline{\boldsymbol{z}} - (\boldsymbol{X} - \boldsymbol{L})^{-1}\underline{\boldsymbol{Z}}\Delta\boldsymbol{x}, \\
\Delta \overline{\boldsymbol{z}} &= (\boldsymbol{X} - \boldsymbol{U})^{-1}\Big(\sigma\mu\boldsymbol{e} - (\boldsymbol{X} - \boldsymbol{U})\overline{\boldsymbol{Z}}\boldsymbol{e} - \overline{\boldsymbol{Z}}\Delta\boldsymbol{x}\Big), \\
&= (\boldsymbol{X} - \boldsymbol{U})^{-1}\sigma\mu\boldsymbol{e} - \overline{\boldsymbol{z}} - (\boldsymbol{X} - \boldsymbol{U})^{-1}\overline{\boldsymbol{Z}}\Delta\boldsymbol{x}).
\end{aligned} \tag{2.1.7}$$

The system Equation (2.1.4) can then be reduced to the following system in $\Delta \boldsymbol{x}$ and $\Delta \boldsymbol{y}$:

$$\begin{aligned}
&\boldsymbol{A}\Delta\boldsymbol{x} = \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}, \\
&-\Big((\boldsymbol{X} - \boldsymbol{L})^{-1}\underline{\boldsymbol{Z}} + (\boldsymbol{X} - \boldsymbol{U})^{-1}\overline{\boldsymbol{Z}}\Big)\Delta\boldsymbol{x} + \boldsymbol{A}^T\Delta\boldsymbol{y} \\
&\quad = \boldsymbol{c} - \boldsymbol{A}^T\boldsymbol{y} - \underline{\boldsymbol{z}} - \overline{\boldsymbol{z}} + \underline{\boldsymbol{z}} + \overline{\boldsymbol{z}} - \sigma\mu(\boldsymbol{X} - \boldsymbol{L})^{-1}\boldsymbol{e} - \sigma\mu(\boldsymbol{X} - \boldsymbol{U})^{-1}\boldsymbol{e}, \\
&\quad = \boldsymbol{c} - \boldsymbol{A}^T\boldsymbol{y} - \sigma\mu((\boldsymbol{X} - \boldsymbol{L})^{-1} + (\boldsymbol{X} - \boldsymbol{U})^{-1})\boldsymbol{e},
\end{aligned} \tag{2.1.8}$$

and $\Delta \underline{\boldsymbol{z}}$ and $\Delta \overline{\boldsymbol{z}}$ are as defined in Equation (2.1.7). Note that this formulation retains the sparsity of the original problem. This allows the usage of sparse linear system solvers, such as Cholesky factorization to efficiently solve the system of equations.

3 Nonlinear Programming algorithms

4 Stochastic Programming algorithms

References