Fast Gradient Method and Dykstra's Alternating Projection

Idris Kempf

February 18, 2022

1 Problem Statement

Consider the state-space representation of a linear system in discrete time,

$$x_{k+1} = Ax_k + Bu_k, \quad u_k \in \mathcal{U}(u_{k-1}), \tag{1a}$$
$$y_k = Cx_k + d_k, \tag{1b}$$

where $y_k \in \mathbb{R}^{n_y}$ are the outputs at time $t = kT_s$ and $x_k \in \mathbb{R}^{n_x}$ are the states. The inputs $u_k \in \mathbb{R}^{n_u}$ are subjected to amplitude and slew-rate constraints that can be modeled as $\mathcal{U}(u_{k-1}) \coloneqq \mathcal{A} \cap \mathcal{R}(u_{k-1})$, where u_{k-1} is the input applied at time $t = (k-1)T_s$ and $\mathcal{A}$ and $\mathcal{R}(u_{k-1})$ are the amplitude and slew-rate constraint sets:

$$\mathcal{A} \coloneqq \{u_k \in \mathbb{R}^{n_u} \mid -\alpha \le u_k \le \alpha\}, \tag{2a}$$
$$\mathcal{R}(u_{k-1}) \coloneqq \{u_k \in \mathbb{R}^{n_u} \mid -\rho \le u_k - u_{k-1} \le \rho\}. \tag{2b}$$

The condensed model predictive control problem for (1a) is

$$\min_u \frac{1}{2} u^{\mathrm{T}} J u + q(\hat{x}_k, \hat{d}_k)^{\mathrm{T}} u \quad \text{s.t.} \quad u \in \mathcal{S}(u_{-1}), \tag{3}$$

where $u \coloneqq (u_0^{\mathrm{T}}, \dots, u_{N-1}^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^{Nn_u}$, N the horizon, $f(u) \coloneqq \frac{1}{2} u^{\mathrm{T}} J u + q(\hat{x}_k, \hat{d}_k)^{\mathrm{T}} u$ the objective function and $J = J^{\mathrm{T}} \in \mathbb{R}^{Nn_u \times Nn_u}$ the Hessian. The vector $q(\hat{x}_k, \hat{d}_k)$ is an affine function of the observer output $\hat{x}_k$ and $\hat{d}_k$. The closed convex set $\mathcal{S}(u_{-1})$ is defined as

$$\mathcal{S}(u_{-1}) \coloneqq \mathcal{U}(u_{-1}) \times \dots \times \mathcal{U}(u_{N-2}), \tag{4}$$

and depends on the input u_{-1} applied at time $t-1$. The following assumptions on problem (3) are made throughout the paper:

Assumption I.a (Strong convexity)**.** *It holds that $0 \prec \lambda_{min} I \preceq J \preceq \lambda_{max} I$, where λ_{min} and λ_{max} are the minimum and maximum eigenvalues of J.*

Assumption I.b (Bounded inputs)**.** *The set $\mathcal{S}$ is bounded, so that $|\mathcal{S}| \coloneqq \max_{x,y \in \mathcal{S}} \|x-y\|_2 < \infty$.*

To solve problem (3), we are using the fast gradient method shown in Algorithm 1. The present variant of the fast gradient method is applicable to strongly convex functions for ill-conditioned Hessians [5, Ch. 2.2.4]. The fast gradient method requires the Eucledian projection $\mathcal{P}_{\mathcal{S}}$ onto $\mathcal{S}$, and no simple formula exists for $N > 1$. Here, we are replacing the exact projection $\mathcal{P}_{\mathcal{S}}$ with Dykstra's alternating projection algorithm, $\mathcal{D}_{\mathcal{S}}$. Dykstra's method is an interative algorithm that yields the exact projection if the algorithm is run for an infinite number of iterations. For M iterations, $\mathcal{D}_{\mathcal{S}}(z) \notin \mathcal{S}$ in general[1] and the method yields a projection error that can be quantified as shown in the following assumption.

[1] For example, consider the projection onto a corner (or edge) of two intersecting hyperplanes.

Assumption II (Approximate projection)**.** *Dykstra's method $\mathcal{D}_{\mathcal{S}}$ returns a point $\mathcal{D}_{\mathcal{S}}(z)$ that satisfies*

$$\|\mathcal{D}_{\mathcal{S}}(z) - \mathcal{P}_{\mathcal{S}}(z)\|_2 \leq \delta(z, M, \mathcal{S}), \tag{5}$$

where the upper bound $\delta(z, M, \mathcal{S}) > 0$ depends on the maximum number of iterations M of Dykstra's method.

The aim of this paper is to

1. Characterize $\delta(z, M, \mathcal{S})$ for $\mathcal{S}$ and $z \in \mathcal{Z}$, where $\mathcal{Z} = \mathcal{Z}(J, q, u_{-1})$ is to be defined.
2. Analyze the converge of Algorithm 1 for fixed M.

The ideal result would be to obtain a formula that relates M and ϵ, where ϵ represents an upper bound on the solution accuracy of the modified fast gradient method.

Algorithm 1 Fast gradient method: Constant step scheme III with parameter [5, Ch. 2.2.4].

Input: $u_{k-1} \to v_0$
Output: $u_k \leftarrow p_M$

```
Set p_0 = 0
for i = 0 to M − 1 do
    z_{i+1} = (I − J λ_max^{-1}) v_i − q λ_max^{-1}
    p_{i+1} = P_S(z_{i+1})
    v_{i+1} = (1 + β) p_{i+1} − β p_i
end for
```

2 Fast Gradient Method

A convergence analysis of Algorithm 2 is presented in [6] with $\mathcal{P}_{\mathcal{S}}$ replaced by $\mathcal{D}_{\mathcal{S}}$ satsifying

$$\|\mathcal{D}_{\mathcal{S}}(z) - \mathcal{P}_{\mathcal{S}}(z)\|_2 \leq \delta, \tag{6}$$

where $\delta > 0$ is an arbitrary constant. Note that Algorithm 1 is obtained from Algorithm 2 by setting $\alpha_0^2 = \lambda_{\max}/\lambda_{\min}$. The authors prove that if

$$\delta \leq \frac{\epsilon \lambda_{\max}^{1/2}}{120 \lambda_{\min}^{3/2} |\mathcal{S}|}, \tag{7}$$

then at most K iterations of Algorithm 2,

$$K = \left\lceil \sqrt{\frac{\lambda_{\max}}{\lambda_{\min}}} \log \left(\frac{3\lambda_{\min} r_1^2 + 6(f(\mathcal{P}_{\mathcal{S}}(z_1)) - f^\star)}{\epsilon} \right) \right\rceil, \tag{8}$$

are required to obtain a solution p_K that satisfies

$$\operatorname{dist}_{\mathcal{S}}(p_K) \leq \epsilon, \qquad f(p_K) - f^\star \leq \epsilon, \tag{9}$$

where $\operatorname{dist}_{\mathcal{S}}(y) \coloneqq \min_{x \in \mathcal{S}} \|x - y\|_2$, $f^\star = f(u^\star)$ is the minimum of (3) and $r_k \coloneqq \|p_k - u^\star\|_2$.

Algorithm 2 Fast gradient method: Constant step scheme III [5, Ch. 2.2.4] with approximate projection.

Input: $u_{k\text{-}1} \to v_0$
Output: $u_k \leftarrow p_M$
Set $p_0 = 0$
for $i = 0$ to $M-1$ **do**
  $z_{i+1} = (I - J\lambda_{max}^{-1})v_i - q\lambda_{max}^{-1}$
  $p_{i+1} = \mathcal{D}_{\mathcal{S}}(z_{i+1})$
  $\alpha_{i+1}^2 = (1-\alpha_{i+1})\alpha_i^2 + \alpha_{i+1}\lambda_{\max}/\lambda_{\min}$
  $\beta_{i+1} = \alpha_i(1-\alpha_i)/(\alpha_i^2 + \alpha_{i+1})$
  $v_{i+1} = (1+\beta_{i+1})p_{i+1} - \beta_{i+1}p_i$
end for

3 Dykstra's Alternating Projection

Given r convex sets $\mathcal{H}_1, \dots, \mathcal{H}_r$, Dykstra's alternating projection algorithm [2, 8] finds the orthogonal projection $x^\star$ of x onto $\mathcal{H} \coloneqq \bigcap_{i=1}^r \mathcal{H}_i$ by iterating over

$$x_n = \mathcal{P}_{\mathcal{H}_{[n]}}\left(x_{n-1} + e_{n-r}\right), \qquad e_n = x_{n-1} + e_{n-r} - x_n, \tag{10}$$

where $[n] = \{n + mr \mid m \in \mathbb{Z}\} \cap \{1, \dots, r\}$, $x_0 = x$ and $e_i = 0$ when $i \leq 0$. The Boyle-Dykstra theorem [2] implies that $\lim_{n\to\infty} \|x_n - \mathcal{P}_{\mathcal{H}}(x)\| = 0$. For a finite number of iterations, there is no guarantee that $x_n \in \mathcal{H}$. Here, we assume that $\mathcal{H}$ is a polyhedron and the $\mathcal{H}_i$ are halfspaces given by

$$\mathcal{H}_i \coloneqq \{x \mid \langle x, f_i\rangle \leq c_i\}, \tag{11}$$

where $\|f_i\| = 1$. In addition, define

$$H_i \coloneqq \{x \mid \langle x, f_i\rangle = c_i\}, \tag{12}$$

so that $\operatorname{int}\mathcal{H}_i \coloneqq \mathcal{H}_i \backslash H_i$. The projections onto H_i and $\mathcal{H}_i$ are given by

$$\mathcal{P}_{H_i}(x) = x - \left(\langle x, f_i\rangle - c_i\right) f_i, \qquad \mathcal{P}_{\mathcal{H}_i}(x) = \begin{cases} x & \text{if } x \in \mathcal{H}_i \\ \mathcal{P}_{H_i}(x) & \text{if } x \notin \mathcal{H}_i\,. \end{cases} \tag{13}$$

For this particular choice of sets, $e_n = k_n f_{[n]}$ with $k_n = \operatorname{dist}_{\mathcal{H}_{[n]}}(x_{n-1} + e_{n-r})$, i.e. the auxiliary vector is either 0 or parallel to $f_{[n]}$.
The convergence of Dykstra's method has been analyzed in [4, 3], where it has been shown that the convergence is linear. In [3], the proof is based on partitioning the set $\{1, \dots, r\}$ into

$$A = \{i \in \{1, \dots, r\} \mid x_\infty \in H_i\}, \qquad B = \{1, \dots, r\} \backslash A = \{i \in \{1, \dots, r\} \mid x_\infty \in \operatorname{int}\mathcal{H}_i\}, \tag{14}$$

where $x_\infty = \lim_{n\to\infty} x_n$. It can be shown that there exists a number N_1 such that whenever

$$[n] \in B, \quad n \geq N_1 \quad \Rightarrow \quad x_n = x_{n-1}, \quad e_n = 0, \tag{15}$$

i.e. the half-spaces in B become "inactive". Furthermore, there exists $N_2 \geq N_1$ such that whenever $n \geq N_2$, it holds that

$$\|x_{n+r} - x_\infty\|_2 \leq \alpha_{[n]} \|x_n - x_\infty\|_2, \tag{16}$$

where $0 \leq \alpha_{[n]} < 1$. With these ingredients, the following result is obtained:

Theorem 1. *There exist constants* $0 \leq c < 1$ *and* $\rho > 0$ *such that*

$$\|x_n - x_\infty\| \leq \rho c^n.$$

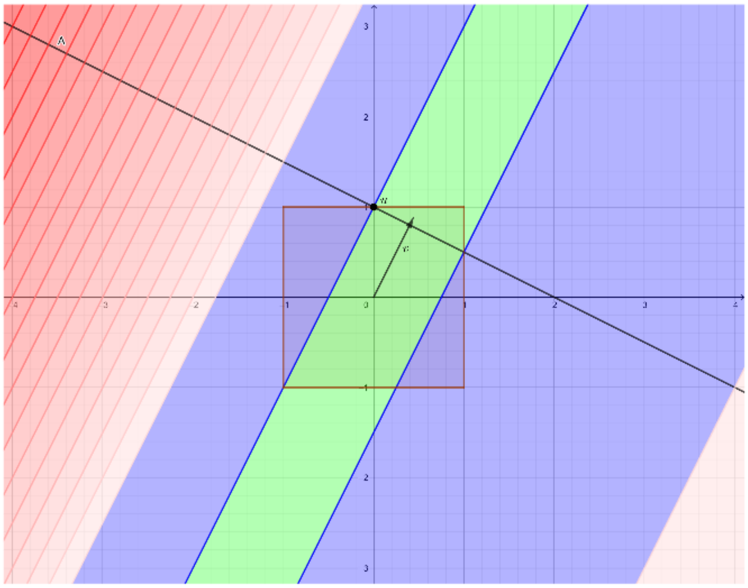

Figure 1: Line-box example with different regions that yield different convergence properties.

The factor c can be estimated from the smallest $\alpha_{[n]}$, which is characterized by the angle between certain subspaces (subspaces formed by the "active" halfspaces). The factor $\alpha_{[n]}$ can be upper-bounded by considering the "worst" angles in the polyhedron.

The constant ρ, however, depends on a number $N_3 \geq N_2$ and on x. It is unclear how to obtain that constant ρ. In fact, the authors of [7] and [9] emphasize that ρ cannot be computed in advance, and that the inability to compute a bound on the projection error makes the application of Dykstra's method difficult. The authors of [7] proposed a combined Dysktra-conjugate-gradient method that allows for computing an upper bound on $\|x_n - x_\infty\|$. The authors of [9] proposed an alternative algorithm called *successive approximate algorithm*, which promises fast convergence, conditioned on knowing a point $x \in \mathcal{H}$ in advance.

4 Stalling

In [1], the behavior of Dykstra's method is analalyzed for two sets. The authors give conditions on Dykstra's algorithm for (1) finite convergence, (2) infinite convergence and (3) stalling followed by infinite convergence. A specific example is given for the case that the set is given by the intersection of a line with a unit box in $\mathbb{R}^2$ ($\mathcal{H}$ is a polyhedron). It can be shown that cases 1–3 depend on the starting point x_0, and one can determine the 3 regions shown in Figure 1 that yield that yield the different convergence behavior. Convergence case 1 is obtained when starting in the green region, case 2 when starting in the blue region and case 3 when starting in the red region.

To understand the stalling effect, consider Figure 2, which shows the first iterations of Dykstra's algorithm with starting point in the red region. Note that the outcome of Dykstra's algorithm depends on the order of the sets $\mathcal{H}_i, \ldots, \mathcal{H}_r$. In Figure 2, the algorithm starts by projecting onto the line and then onto the box. It can be seen that for the first 6 iterations[2], Dykstra's algorithm returns the top left corner of the box ("stalling"). The author's also determine the exact number of iterations required to break-free from the red region, and show that if the starting point is arbitrarily far to the left, the algorithm will need an arbitrarily large iteration number for breaking-free from the red region.

[2]By one iteration we mean one cycle of r projections here.

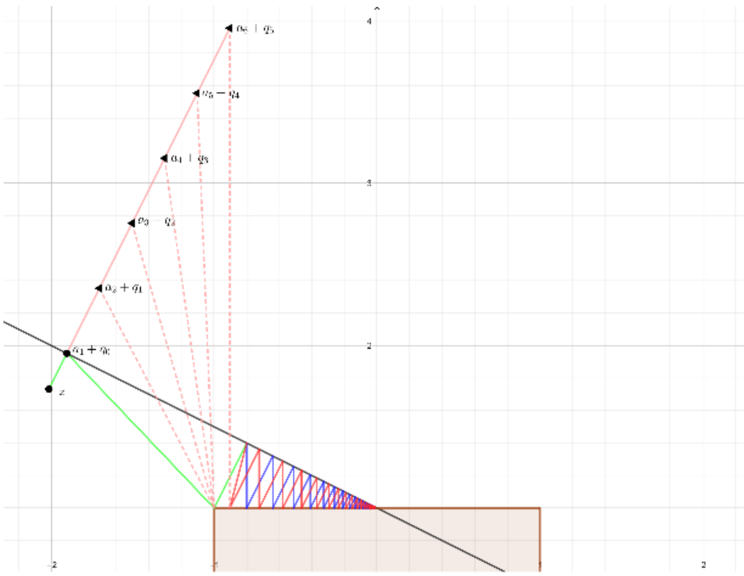

Figure 2: Stalling for the line-box example when x_0 is in the red region.

5 Thoughts

In theory, the formulae presented in these notes would allow for determining the solution accuracy of the fast gradient method if both the fast gradient method and Dykstra's algorithm are run for a fixed number of iterations. However, because the error bound for Dykstra's method cannot easily be computed in advance[3], the formulae cannot be applied.
The results on the fast gradient method are surprising. I do not understand how one can obtain a solution with ϵ accuracy if the approximate projection returns the unchanged vector, for example. However, it does seem that the proof requires the projection error to be bounded by δ (regardless of the starting point), in which case the former example does not work.
It seems obvious that there is a connection between the stalling and the inability to compute the error bound in advance. E.g. consider the stalling situation from last section (Figure 2). It is my impression that the constant c in the Dykstra error bound could be computed after the stalling period (note that N_2 is reached after the stalling period – one needs to model the box via halfplanes, in which case we enter the interior of the halfplane defining the left side of the box).
Questions:

- Can Dysktra's method be modified to escape the stalling region in a finite number of iterations?
  It does seem that introducing $e_n = e_{n-r} + \beta_n(x_{n-1} - x_n)$ improves things.
- Can Dysktra's method be accelerated after escaping the stalling region?
- Can we replace Dykstra's method by the *successive approximate algorithm* [9]?
  The previously computed input is always feasible.
- Can we modify Dykstra's method such that it becomes independent of the set ordering?
  This should be possible to formulate mathematically. However, it does make sense to obtain different "trajectories". The convergence properties should be independent of set ordering.

[3] For an arbitrary polyhedral set. It can be computed for e.g. 2 sets.

- Is there some heuristic for the set ordering?
  I have tried a few without success.

6 Acceleration for Dykstra's Method

Consider introducing the step size parapemeter $\beta_n \geq 0$:

$$x_n = \mathcal{P}_{\mathcal{H}_{[n]}}\left(x_{n-1} + e_{n-r}\right), \qquad e_n = e_{n-r} + \beta_n(x_{n-1} - x_n). \tag{17}$$

For $\beta_n = 0$, we obtain Von Neumann's *Alternating Projection Method* and for $\beta_n = 1$, we obtain Dykstra's method (10). We proceed by characterizing the term e_n.

Lemma 1. *It holds that $e_n = y_n f_{[n]}$, where*

$$y_n = (1 - \beta_n) y_{n-r} + k_n,$$

and $k_n = dist_{\mathcal{H}_{[n]}}(x_{n-1} + e_{n-r})$.

Proof. Suppose that $x_{n-1} + e_{n-r} \in \operatorname{int} \mathcal{H}_{[n]}$. Then, $x_n = x_{n-1} + e_{n-r}$ and

$$e_n = (1 - \beta_n) e_{n-r}.$$

Suppose that $x_{n-1} + e_{n-r} \notin \operatorname{int} \mathcal{H}_{[n]}$. Then, $x_n = x_{n-1} + e_{n-r} - \left(\left\langle x_{n-1} + e_{n-r}, f_{[n]}\right\rangle - c_i\right) f_{[n]}$ and

$$e_n = (1 - \beta_n) e_{n-r} + k_n f_{[n]}.$$

Note that $e_n = 0$ for $n \leq 0$. By induction, e_n is always parallel to $f_{[n]}$ or zero. Substituting $e_n = y_n f_{[n]}$ yields

$$y_n = (1 - \beta_n) y_{n-r} + k_n.$$

$\square$

References

[1] H. H. Bauschke, R. S. Burachik, D. B. Herman, and C. Y. Kaya. On Dykstras algorithm: finite convergence, stalling, and the method of alternating projections. *ArXiv e-prints*, pages 1–14, January 2020.

[2] J. P. Boyle and R. L. Dykstra. A method for finding projections onto the intersection of convex sets in hilbert spaces. In *Advances in Order Restricted Statistical Inference*, pages 28–47. Springer New York, 1986.

[3] F. Deutsch and H. Hundal. The rate of convergence of Dykstra's cyclic projections algorithm: The polyhedral case. *Numerical Functional Analysis and Optimization*, 15(6):537–565, May 1994.

[4] A. N. Iusem and A. R. de Pierro. On the convergence properties of Hildreth's quadratic programming algorithm. *Math. Prog. Ser. A and B*, 47(1):37–51, May 1990.

[5] Y. Nesterov. *Introductory Lectures on Convex Optimization: A Basic Course*. Applied Optimization. Springer Science & Business Media, 2003.

[6] A. Patrascu and I. Necoara. On the convergence of inexact projection primal first-order methods for convex minimization. *IEEE Trans. on Automat. Cont.*, 63(10):3317–3329, Oct. 2018.

[7] C. Perkins. A convergence analysis of Dykstra's algorithm for polyhedral sets. *SIAM J. Numer. Anal.*, 40(2):732–804, July 2002.

[8] H. Shih-Ping. A successive projection method. *Math. Prog.*, 40(1):1–14, January 1988.

[9] S. Xu. Successive approximate algorithm for best approximation from a polyhedron. *J. Approx. Theory*, 93(3):415–433, June 1998.