

# DeePC and DRO Notes

Alessandro Quattrociochi

December 4, 2024

## Contents

<b>1</b>	<b>Problem Formulation</b>	<b>1</b>
1.1	Conventional Formulation . . . . .	1
1.2	Deterministic DeePC Formulation . . . . .	2
1.2.1	Remark I: <i>DeepC representation of predictor.</i> . . . .	2
1.2.2	Remark: <i>Regularization in DeePC(rank-deficient Hankel matrices)</i> . . . . .	2
<b>2</b>	<b>Stochastic DeePC Formulation</b>	<b>3</b>
2.1	Wasserstein Distance . . . . .	4
2.2	Objective Function Reformulation . . . . .	4
2.3	Constraint Reformulation . . . . .	5
2.4	DRO Chance Constrained DeePC- Tractable Reformulation . . . . .	5

## 1 Problem Formulation

Consider a deterministic LTI system whose model is unknown. Let  $T_f \in \mathbb{Z}_{>0}$  be the prediction horizon. Let  $f_1 : \mathbb{R}^{mT_f} \rightarrow \mathbb{R}_{\geq 0}$ , and  $f_2 : \mathbb{R}^{pT_f} \rightarrow \mathbb{R}_{\geq 0}$  be a cost function on the future inputs and, respectively, outputs. We aim at finding a input sequece  $\text{col}(u_t, \dots, u_{t+T_f-1}) \in \mathbb{R}^{mT_f}$ , such that  $\text{col}(y_t, \dots, y_{t+T_f-1}) \in \mathbb{R}^{pT_f}$  minimizes the cost  $f_1 + f_2$ , and the constraints satisfied, i.e.,  $u \in \mathcal{U}$  and  $y \in \mathcal{Y}$ , with  $\mathcal{U} \subseteq \mathbb{R}^{mT_f}$  and  $\mathcal{Y} \subseteq \mathbb{R}^{pT_f}$ .

### 1.1 Conventional Formulation

Using conventional input output state space representation.

$$\begin{aligned} \min_{u,y,x} \quad & f_1(u) + f_2(y) \\ \text{s.t.} \quad & \forall k \in \{0, \dots, T_f - 1\} \\ & x_{k+1} = Ax_k + Bu_k \\ & y_k = Cx_k + Du_k \\ & x_0 = \hat{x}_t \\ & u \in \mathcal{U}, y \in \mathcal{Y} \end{aligned} \tag{1}$$

## 1.2 Deterministic DeePC Formulation

From data we construct  $\hat{U}_p, \hat{U}_f, \hat{Y}_p, \hat{Y}_f$ .

$$\begin{aligned} \min_g \quad & f_1(u) + f_2(y) \\ \text{s.t.} \quad & \begin{bmatrix} \hat{U}_p \\ \hat{Y}_p \\ \hat{U}_f \\ \hat{Y}_f \end{bmatrix} g = \begin{bmatrix} u_{\text{ini}} \\ y_{\text{ini}} \\ u \\ y \end{bmatrix} \\ & u \in \mathcal{U}, y \in \mathcal{Y} \end{aligned} \tag{2}$$

- $U_p, Y_p$  represent the past input-output trajectories.
- $U_f, Y_f$  represent the future input-output trajectories.

### 1.2.1 Remark I: *DeepC representation of predictor.*

$$\begin{bmatrix} u_{\text{ini}} \\ y_{\text{ini}} \\ u \\ y \end{bmatrix} = H(w)g, \quad \text{with} \quad H(w) = \begin{bmatrix} U_p \\ Y_p \\ U_f \\ Y_f \end{bmatrix} \tag{3}$$

where  $H(w)$  is the suitably partitioned Hankel matrix, or

$$y = Y_f g \tag{4}$$

where  $g$  is computed from:

$$\begin{bmatrix} U_p \\ Y_p \\ U_f \end{bmatrix} g = \begin{bmatrix} u_{\text{ini}} \\ y_{\text{ini}} \\ u \end{bmatrix}.$$

The explicit solution for  $y$  is the superposition of a particular solution and the homogenous term:

$$y = Y_f \cdot \left( \begin{bmatrix} U_p \\ Y_p \\ U_f \end{bmatrix}^+ \begin{bmatrix} u_{\text{ini}} \\ y_{\text{ini}} \\ u \end{bmatrix} + \ker \left( \begin{bmatrix} U_p \\ Y_p \\ U_f \end{bmatrix} \right) \right),$$

where  $\begin{bmatrix} U_p \\ Y_p \\ U_f \end{bmatrix}^+$  is the pseudoinverse of the Hankel matrix, and  $\ker$  represents the kernel of the matrix.

### 1.2.2 Remark: *Regularization in DeePC(rank-deficient Hankel matrices)*

When the Hankel matrix  $H$  is rank-deficient (i.e.,  $H$  has more columns than rows), there are infinitely many solutions  $g$  that satisfy the system equation:

$$Hg = \begin{bmatrix} u_{\text{ini}} \\ y_{\text{ini}} \\ u \end{bmatrix}. \tag{5}$$

To find a noise-robust prediction, the Projection Regularizer is introduced. The regularizer penalizes the portion of  $g$  that lies in the nullspace of  $H$ . This is achieved by minimizing:

$$\|(I - \Pi)g\|_2^2, \quad (6)$$

where:

- $\Pi$  is the orthogonal projection matrix onto the row space of  $H$ ,
- $(I - \Pi)g$  isolates the component of  $g$  in  $\ker(H)$ , the nullspace of  $H$ .

This ensures that the computed  $g$  is the smallest norm solution to:

$$y = Y_f g. \quad (7)$$

In the noiseless case ( $\lambda_s = 0$ ) with regularization constant  $\lambda_p \geq 0$ , the optimization problem ensures that both  $u$  and  $y$  match the reference trajectory  $u_r, y_r$ , making it beneficial to use the regularizer with a sufficiently large  $\lambda_p$ .

## 2 Stochastic DeePC Formulation

Following the formulation in [2], we now extend the DeePC definition of the unknown LTI system by considering a disturbance vector  $w_t \in \mathbb{R}^q$ , such as:

$$\begin{cases} x_{t+1} = Ax_t + Bu_t + Ew_t \\ y_t = Cx_t + Du_t + Fw_t \end{cases} \quad (8)$$

where  $E \in \mathbb{R}^{n \times q}$  and  $F \in \mathbb{R}^{p \times q}$ . Therefore, the system is subject to an **unknown** and **uncontrollable** disturbance, whose past trajectory can be measured but whose future evolutions are unknown. Similar to  $u^d$  and  $y^d$ , we can build the Hankel matrix of the disturbance  $\mathcal{H}_{T_{ini+N}}$ , splitted as follows:

$$\begin{bmatrix} W_P \\ W_F \end{bmatrix} := \mathcal{H}_{T_{ini+N}}(w^d) \quad (14)$$

with  $W_P \in \mathbb{R}^{qT_{ini} \times H_c}$  and  $W_F \in \mathbb{R}^{qN \times H_c}$ . Therefore, we assume exist  $g \in \mathbb{R}^{H_c}$ , so that:

$$\begin{bmatrix} U_P \\ W_P \\ Y_P \\ U_F \\ W_F \\ Y_F \end{bmatrix} g = \begin{bmatrix} u_{ini} \\ w_{ini} \\ y_{ini} \\ u \\ w \\ y \end{bmatrix} \quad (9)$$

Unlike [2], we DO NOT assume  $w_t \in [\underline{w}, \bar{w}]^q$ . Why this? We assume that in optimal scheduling problems, a time series forecasting model is always subject to an error which in very few toy examples converges to white noise. Therefore, we assume the general case where the error term distribution  $w_t$  is not known a priori. Therefore, we assume  $\hat{W}_P$

and  $\hat{W}_F$  realizations of random variables  $W_P$  and  $W_F$ . Reformulating from [1], we define  $\xi = (\xi_1^T, \dots, \xi_{p, (T_{\text{ini}} + T_f)})$ , with  $\xi_i$  denoting the  $i$ th rows ( $W_P^T, W_F^T$ ).

### Objective Function

The objective function formulated in Eq.1, is therefore reformulated into its robust counterpart, namely:

$$\begin{aligned} \min_g \sup_{\mathbb{Q} \in \hat{\mathcal{P}}} \mathbb{E}_{\mathbb{Q}} \left[ f_1(\hat{U}_f g) + f_2(Y_f g) \right] \\ \text{s.t.} \quad \begin{bmatrix} U_P \\ W_P \\ Y_P \\ U_F \\ W_F \\ Y_F \end{bmatrix} g = \begin{bmatrix} u_{\text{ini}} \\ w_{\text{ini}} \\ y_{\text{ini}} \\ u \\ w \\ y \end{bmatrix} \\ \sup_{\mathbb{Q} \in \hat{\mathcal{P}}} \text{CVaR}_{1-\alpha}^{\mathbb{Q}}(h(y)) \leq 0. \end{aligned} \quad (10)$$

## 2.1 Wasserstein Distance

Definition copied from [1],

We now introduce the *Wasserstein ambiguity set* centred around the so-called *empirical distribution*, i.e.,  $\hat{\mathbb{P}} = \frac{1}{N} \sum_{i=1}^N \delta_{\xi}$  [3] where  $\delta_{\xi}$  is the kroneker delta, poiting a discrete sampling from the true unknow distribution.

$\mathcal{M}(\Xi)$  being the set of all distributions  $\mathbb{Q}$  supported on  $\Xi$ , such that  $\mathbb{E}_{\mathbb{Q}}[\|\xi\|_r] < \infty$ , where  $\|\cdot\|_r$  is the  $r$ -norm for some  $r \in [1, \infty]$ .

The *Wasserstein metric*  $d_W : \mathcal{M}(\Xi) \times \mathcal{M}(\Xi) \rightarrow \mathbb{R}_{\geq 0}$  is defined as

$$d_W(\mathbb{Q}_1, \mathbb{Q}_2) := \inf_{\Pi} \left\{ \int_{\Xi^2} \|\xi_1 - \xi_2\|_r \Pi(d\xi_1, d\xi_2) \right\} \quad (11)$$

where  $\Pi$  is a joint distribution of  $\xi_1$  and  $\xi_2$  with marginal distributions  $\mathbb{Q}_1 \in \mathcal{M}(\Xi)$  and  $\mathbb{Q}_2 \in \mathcal{M}(\Xi)$ , respectively. The Wasserstein metric can be viewed as a distance between probability distributions, where the distance is calculated via an optimal mass transport plan  $\Pi$ .

## 2.2 Objective Function Reformulation

By using the distributionally robust linear regression theorem [Theorem 4, II] [4], the objective function can be reformulated as follows:

$$\min_g \sup_{\mathbb{Q} \in \hat{\mathcal{P}}} \mathbb{E}_{\mathbb{Q}} \left[ f_1(\hat{U}_f g) + f_2(Y_f g) \right] = \min_{g, \tau, s_i} f_1(\hat{U}_f g) + \frac{1}{N} \sum_{i=1}^N f_2(\hat{Y}_P^{(i)} g) + L_{\text{obj}} \epsilon \|g\|_{r,*} \quad (12)$$

## 2.3 Constraint Reformulation

$$\sup_{\mathbb{Q} \in B_\epsilon(\hat{\mathbb{P}}_N)} \text{CVaR}_{1-\alpha}^{\mathbb{Q}}(h(W_f g)) \leq 0 \rightarrow \sup_{\mathbb{Q} \in B_\epsilon(\hat{\mathbb{P}}_N)} \inf_{\tau \in \mathbb{R}} -\tau\alpha + \mathbb{E}_{\mathbb{Q}}[(\ell(g, \xi) + \tau)_+] \leq 0 \quad (13)$$

$$-\tau\alpha + L_{\text{con}}\epsilon\|g\|_{r,*} + \frac{1}{N} \sum_{i=1}^N s_i \leq 0 \quad (14)$$

$$\tau + h(\hat{W}_f^{(i)} g) \leq s_i \quad \forall i \leq N \quad (15)$$

$$s_i \geq 0 \quad \forall i \leq N \quad (16)$$

## 2.4 DRO Chance Constrained DeePC- Tractable Reformulation

$$\min_{g, \tau, s_i} f_1(\hat{U}_f g) + \frac{1}{N} \sum_{i=1}^N \left( f_2(\hat{Y}_f^{(i)} g) \right) + L_{\text{obj}}\epsilon\|g\|_{r,*} \quad (17)$$

$$\text{s.t.} \quad \hat{U}_P g = \hat{u}_{\text{ini}} \quad (18)$$

$$\hat{Y}_P g = \hat{y}_{\text{ini}} \quad (19)$$

$$\hat{U}_f g \in \mathcal{U} \quad (20)$$

$$\hat{Y}_f g \in \mathcal{Y} \quad (21)$$

$$-\tau\alpha + L_{\text{con}}\epsilon\|g\|_{r,*} + \frac{1}{N} \sum_{i=1}^N s_i \leq 0 \quad (22)$$

$$\tau + h(\hat{W}_f^{(i)} g) \leq s_i \quad (23)$$

$$s_i \geq 0 \quad \forall i \leq N \quad (24)$$

# Appendix

By definition of CVaR

$$\text{CVaR}_{1-\alpha}^{\mathbb{P}}(h(y)) \leq 0 \iff \inf_{\tau \in \mathbb{R}} -\tau\alpha + \mathbb{E}_{\mathbb{P}}[(h(y) + \tau)_+] \leq 0. \quad (25)$$

This can be shown by multiplying the definition of CVaR by  $\alpha > 0$  and changing  $\tau$  to  $-\tau$ . Letting  $\ell(g, \xi) := h(\xi_{p(T_{\text{ini}}+1)}^{\top} g, \dots, \xi_{p(T_{\text{ini}}+T_f)}^{\top} g)$ , the constraint  $g \in G_{\text{CVaR}}$  is equivalent to  $g$  satisfying

$$\sup_{\mathbb{Q} \in B_{\epsilon}(\hat{\mathbb{P}}_N)} \inf_{\tau \in \mathbb{R}} -\tau\alpha + \mathbb{E}_{\mathbb{Q}}[(\ell(g, \xi) + \tau)_+] \leq 0 \quad (26)$$

We have

$$\sup_{\mathbb{Q} \in B_{\epsilon}(\hat{\mathbb{P}}_N)} \inf_{\tau \in \mathbb{R}} -\tau\alpha + \mathbb{E}_{\mathbb{Q}}[(\ell(g, \xi) + \tau)_+] \leq \inf_{\tau \in \mathbb{R}} -\tau\alpha + \sup_{\mathbb{Q} \in B_{\epsilon}(\hat{\mathbb{P}}_N)} \mathbb{E}_{\mathbb{Q}}[(\ell(g, \xi) + \tau)_+] \quad (27)$$

by the max-min inequality. From Lemma A.1 we have

$$\sup_{\mathbb{Q} \in B_{\epsilon}(\hat{\mathbb{P}}_N)} \mathbb{E}_{\mathbb{Q}}[(\ell(g, \xi) + \tau)_+] \quad (28)$$

$$= \inf_{\lambda \geq 0} \lambda\epsilon + \frac{1}{N} \sum_{i=1}^N \sup_{\xi \in \Xi} ((\ell(g, \xi) + \tau)_+ - \lambda \|\xi - \hat{\xi}^{(i)}\|_r) \quad (29)$$

Recalling that  $\ell(g, \xi) = h(\xi_{p(T_{\text{ini}}+1)}^{\top} g, \dots, \xi_{p(T_{\text{ini}}+T_f)}^{\top} g)$  and using Lemma A.2 gives

$$\sup_{\mathbb{Q} \in B_{\epsilon}(\hat{\mathbb{P}}_N)} \mathbb{E}_{\mathbb{Q}}[(\ell(g, \xi) + \tau)_+] \leq \begin{cases} \inf_{\lambda \geq 0} \lambda\epsilon + \frac{1}{N} \sum_{i=1}^N (\ell(g, \hat{\xi}^{(i)}) + \tau)_+ & \text{if } L_{\text{con}} \|g\|_{r,*} \leq \lambda \\ \infty, & \text{otherwise.} \end{cases} \quad (30)$$

Hence, by resolving the inf and resorting to an epigraph formulation, we arrive at

$$\inf_{\tau \in \mathbb{R}} -\tau\alpha + \sup_{\mathbb{Q} \in B_{\epsilon}(\hat{\mathbb{P}}_N)} \mathbb{E}_{\mathbb{Q}}[(\ell(g, \xi) + \tau)_+] \quad (31)$$

$$\leq \begin{cases} \inf_{\tau \in \mathbb{R}, s_i \in \mathbb{R}} -\tau\alpha + \epsilon L_{\text{con}} \|g\|_{r,*} + \frac{1}{N} \sum_{i=1}^N s_i \\ \text{s.t. } \tau + \ell(g, \hat{\xi}^{(i)}) \leq s_i \\ s_i \geq 0 \\ \forall i \leq N \end{cases} \quad (32)$$

Substituting the definition for  $\ell$  and recalling from (13) that we wish for the right hand side of the above inequality to be nonnegative gives

$$\begin{cases} \inf_{\tau, s_i} -\tau\alpha + L_{\text{con}}\epsilon \|g\|_{r,*} + \frac{1}{N} \sum_{i=1}^N s_i \leq 0 \\ g \mid \tau + h(\hat{Y}_f^{(i)} g) \leq s_i \\ s_i \geq 0 \\ \forall i \leq N \end{cases} \subseteq G_{\text{CVaR}}. \quad (33)$$

---

## References

- [1] Jeremy Coulson, John Lygeros, and Florian Dörfler. Distributionally robust chance constrained data-enabled predictive control. *IEEE Transactions on Automatic Control*, 67(7):3289–3304, 2021.
- [2] Linbin Huang, Jeremy Coulson, John Lygeros, and Florian Dörfler. Decentralized data-enabled predictive control for power system oscillation damping. *IEEE Transactions on Control Systems Technology*, 30(3):1065–1077, 2021.
- [3] Peyman Mohajerin Esfahani and Daniel Kuhn. Data-driven distributionally robust optimization using the wasserstein metric: Performance guarantees and tractable reformulations. *Mathematical Programming*, 171(1):115–166, 2018.
- [4] Soroosh Shafieezadeh-Abadeh, Daniel Kuhn, and Peyman Mohajerin Esfahani. Regularization via mass transportation. *Journal of Machine Learning Research*, 20(103):1–68, 2019.