Real-time Optimal Resource Allocation using Online Primal Decomposition

Risvan Dirza∗ Md Rizwan∗∗ Sigurd Skogestad∗ Dinesh Krishnamoorthy∗

∗ Dept. of Chemical Engineering, Norwegian Univ. of Science & Technology (NTNU), NO-7491 Trondheim, Norway, (e-mail: risvan.dirza@ntnu.no; sigurd.skogestad@ntnu.no; dinesh.krishnamoorthy@ntnu.no (corresp. author)).

∗∗ Dept. of Mechanical and Industrial Engineering, Norwegian Univ. of Science & Technology (NTNU), NO-7491 Trondheim, Norway, (e-mail: md.rizwan@ntnu.no).

Abstract:
Many processes in the industry often consist of several subsystems (i.e., clusters) that share common constraints. Typically, each subsystem strives for its objective by competing in obtaining the shared resource, e.g., raw materials extracted from mining activities and production or processing plant capacity. A distributed optimization can solve such a problem, however, it involves solving a numerical optimization problem online and is usually computationally extensive. One can utilize online iteration of Dual decomposition (without numerical solver) to solve such a problem. However, in this approach the constraint is typically controlled on a slow time scale causing significant dynamic constraint violation in the transient, especially in active constraint region switching. In practice, a "back-off" strategy is necessary, and it may lead to profit loss in the long run. To address this issue, we propose to utilize online Primal decomposition instead, where the problem turns into a feedback-based problem, and the constraint controller(s) distribute local setpoints without violating the common constraint. The simulation results show that the proposed approach can reach the ideal steady-state optimum.

Keywords: Distributed optimization, Primal decomposition, Feedback control, Production optimization

1. INTRODUCTION

In recent times process industries are committed to reducing their environmental footprints to ensure long-term sustainable production and thereby tackle climate change. This action includes efforts to improve the efficiency of energy and resource (e.g., raw materials extracted from mining activities, and production or processing plant capacity) of the processes during operation, which can be achieved by implementing real-time optimization (RTO). For the most part, the scope of RTO, in general, was restricted to simple tools for unit operations or small-scale processes within a larger site. In large-scale systems, this takes shape as a decentralized RTO structure where some clusters of the operating units are optimized while the system-wide optimal operation is not achieved.

One potential solution to achieve system-wide optimal operation is distributed optimization framework (Wenzel et al. (2016)). In this framework, the large-scale problem is decomposed into several smaller sub-problems and a central problem coordinates these local subsystems to achieve global optimality. The different strategies for decomposing these large-scale problems can be broadly classified as primal decomposition and dual decomposition methods (Boyd et al. (2008)). In primal decomposition methods, the central problem allocates existing shared resources by directly giving each subproblem the number of resources that it can use, hence these methods are sometimes referred to as auction-based algorithms. In dual decomposition, usually referred to as price-based methods, the central coordinator sets the price for the resources to each subproblem, this marked price drives the decision on the number of resources that the subproblem will use. To this end in both primal and dual decomposition, the different subproblems and the central problem are solved online iteratively, until the problem converges to a feasible and optimal solution.

Decomposition strategies are an active field of research in both real-time optimizations (RTO) as well as model predictive control (MPC) (Maestre et al. (2014)). These strategies have several advantages, i.e., allowing distributed implementation (especially for a problem with additively separable cost), and explicit constraint control. However, their adoption in practice requires solving a numerical optimization problem online which is a fundamental limiting factor for process industries due to issues related to numerical robustness and computational
cost. Additionally, the expected benefits of a numerical optimization-based RTO are best realized with regular maintenance and monitoring (Shook (2006)), which requires expert knowledge. For these reasons, the implementation of a distributed RTO in large-scale process systems with complex interconnections has been limited. Lastly, and more importantly, it allows for a separation in time scale of the subsystems (because of distributed implementation).

To this end, there has been increased interest in the development of feedback-optimizing control. In this approach, the economic objectives are translated into control objectives and the optimal plant operation is achieved by directly manipulating inputs using the feedback. This concept was introduced in the 1980s by Morari et al. (1980). Since then different feedback-based real-time optimization (RTO) methods have been developed further (Engell, 2007; Jäschke et al., 2017).

Recent articles such as Krishnamoorthy (2021) and Dirza et al. (2021) proposed a distributed RTO framework based on a central price coordinator scheme (i.e., dual decomposition) that uses simple feedback control tools to ensure that the closed-loop trajectories of the different subsystems converge to a stationary solution of the system-wide optimization problem. This proposed framework was limiting in the sense that the coupling constraint may not be feasible during the transients. Hence, the primal feasibility was guaranteed only upon the convergence to the system-wide steady-state optimal solution. This would require introducing back-off and accepting associated loss to address primal infeasibility during transients.

In this work, we aim to address the issue of primal infeasibility and propose an online optimization method for optimal resource sharing based on primal decomposition using simple feedback controllers. The proposed framework is especially attractive wherein the associated subsystems share their local Lagrange multiplier of the shared resources (also known as local shadow price).

The main contribution of this paper is a distributed feedback-based real-time optimization framework based on primal decomposition, that achieves optimal steady-state operation in a distributed manner, without the need to solve numerical optimization problems online and with minimum dynamic constraint violation.

We demonstrate the proposed method in a network of gas-lifted oil wells production system. These oil wells are operated locally and share a common processing facility at the topside. Using recent technology, the subsea production wells have capabilities for the measurement of the multi-phase flow rates (i.e. multiphase flowmeter, MPFM or virtual flowmeter, VFM technology solutions) at respective wellheads (Hansen et al., 2019). Since the produced gas handling capacity available on a platform is usually limited, it is necessary to optimally allocate the lift gas among the different wells.

The remainder of the paper is organized as follows. Section 2 describes the optimization problem formulation for the overall production facility. Section 3 describes the proposed method. Section 4 presents the simulation example of gas-lifted oil production optimization with limited produced gas handling capacity before concluding the paper in Section 5.

2. PROBLEM FORMULATION

In this section, we describe the optimization problem for the entire system consisting of a network of $N$ subsystems. These subsystems are denoted by the set $\mathcal{N} = \{1, \ldots, N\}$. Due to practical reasons, i.e., it is easier to operate a small system, we assume each subsystem is optimized locally.

Let subsystem $i$ be modeled as a nonlinear state-space system.

$$
\begin{align*}
\dot{x}_i &= f_i(x_i, u_i, d_i) \\
y_i &= h_i(x_i, u_i, d_i)
\end{align*}
$$

where $x_i \in \mathbb{R}^{n_{xi}}$, $u_i \in \mathbb{R}^{n_{ui}}$, $d_i \in \mathbb{R}^{n_{di}}$ and $y_i \in \mathbb{R}^{n_{yi}}$ denote the vector of states, inputs, disturbances/parameters and available measurements of each subsystem, respectively. Each subsystem may also have local constraints.

We consider the overall network as a nonlinear state-space system and define all inputs, states, and disturbances as shown in the following.

$$
\begin{align*}
\mathbf{u} &= [u_1, \ldots, u_N]^\top \\
\mathbf{x} &= [x_1, \ldots, x_N]^\top \\
\mathbf{d} &= [d_1, \ldots, d_N]^\top
\end{align*}
$$

(2)

The steady-state optimization problem is

$$
\begin{align*}
\min_{\mathbf{u}, \mathbf{y} \in \mathcal{N}} & \quad J_N = \sum_{i \in \mathcal{N}} J_{N_i} \\
\text{s.t.} & \quad f(x_i, u_i, d_i) = 0, \\
& \quad g(x_i, u_i, d_i) \leq 0
\end{align*}
$$

(3a)

where constraint (3b) is related to the entire system model, and constraint (3c) is a (coupling) (in-)equality constraint.

3. DISTRIBUTED FEEDBACK-OPTIMIZING CONTROL USING ONLINE PRIMAL DECOMPOSITION

3.1 Distributed Optimization using Primal Decomposition

Solving the integrated optimization problem (3) requires a detailed model and their interactions in addition to the constraints and measurements, which may be undesirable or unnecessary in the practical context. Therefore, we propose to solve problem (3) in a distributed manner by decomposing the problem. In this paper, we propose an online optimization method, i.e., using simple feedback controllers, based on primal decomposition and addressing the issue of primal infeasibility.

First, we introduce a virtual subsystem denoted as subsystem 0, in which the cost function is $J_{N,0} = 0$. As a consequence, we define the set $\mathcal{N}_0 = \{0, 1, \ldots, N\}$.

Defining constraint (3c) as linear constraint (if it is non-linear, one can consider to linearize it at the operating point), $g(x, u, d) = \sum_{i=0}^N g_i(x_i, u_i, d_i) - g_{\text{max}}$, we import a slack variable, $g_0$, to convert any inequality constraint in (3c) into equality constraints, where $g(x, u, d) + g_0 = 0$. This modification does not change the structure that (3a) is additively separable in the cost, and the system model (3b) are imposed for each subsystem independently.

By providing an initial value of local constraint for the variables of the coupling constraint, labeled by $g_0^{\text{sp}}$, where

$$
\begin{align*}
\mathbf{g} &= [g_1, \ldots, g_N]^\top \\
\mathbf{g}_{\text{sp}} &= [g_0^{\text{sp}}, \ldots, g_N^{\text{sp}}]^\top
\end{align*}
$$

(4)

The resulting optimization problem is

$$
\begin{align*}
\min_{\mathbf{u}, \mathbf{y} \in \mathcal{N}} & \quad J_N = \sum_{i \in \mathcal{N}} J_{N_i} \\
\text{s.t.} & \quad f(x_i, u_i, d_i) = 0, \\
& \quad g(x_i, u_i, d_i) + g_0 = 0
\end{align*}
$$

(5)

where constraint (5a) is related to the entire system model, and constraint (5c) is a (coupling) (in-)equality constraint.
where we may consider an integrating controllers with integral gain $K_{I,i} = rac{1}{(τ_{L,i})(τ_{s,i})}$, and $K_i$ is the step response gain, and $τ_{s,i}$ is the desired closed-loop time constant. Note that the desired time constant should be slow enough to satisfy the time-scale separation concept (Baldea and Daoutidis, 2007). This concept is necessary to avoid undesired behaviors such as oscillatory and deviating behavior.

Note that to compensate any change in the normal subsystem, we assume that each subsystem informs its local Lagrange multipliers $λ^i_0$ to the compensator subsystem, and receive the local Lagrange multipliers $λ^c_N$ of the compensator subsystem.

Virtual subsystem: Since we introduce a slack variable $g^{0p}_{sp}$ to store un-utilized resource, and the storage is physically never been negative, it is necessary to use max selector as follows.

$$g^{0p}_{sp,k+1} = \max \left[ 0, g^{0p}_{sp,k} + K_{I,0} \nabla g^{0p}_{sp} \left( \sum_{i \in N_0} P_i (g_{sp,k}^i) \right) \right]$$ (10)

By implementing these strategies, i.e., compensator, normal, and virtual subsystem, the setpoints, provided by these controllers, guarantee the primal feasibility.

Local setpoint controllers: Given the local setpoint $g^{sp}_{sp}$, the local setpoint controller regulates the actual local primal variables $g_i$ to $g^{sp}_{sp}$. The updated local input at the next time step $u^{sp,k+1}_i$ is given by

$$u^{sp,k+1}_i = u^{sp,k}_i + K_{I,i} \left( g_i - g^{sp}_{sp} \right)$$ (11)

where we may consider an integrating controllers with integral gain $K_{I,i} = \frac{1}{(τ_{L,i})(τ_{s,i})}$, and $K_i$ is the step response gain, and $τ_{s,i}$ is the desired closed-loop time constant. Typically, the desired time constant is designed as fast as possible. However, it is necessary to carefully choose the desired time constant $τ_{s,i}$ to ensure that the local setpoint controller does not too aggressively track the setpoint given by central constraint controllers.

Remarks: Note that the setpoint controller is not necessary when we have a shared input constraint because the central constraint controller has provided the optimal input.

Local Lagrange Multiplier estimation: Subgradient (8) is the evaluation of local Lagrange multipliers from each subproblem. According to KKT (Karush-Kuhn-Tucker) conditions, the stationary point is reached when

$$\nabla u_i L_i (u_i, g_{sp}, λ_i) = 0$$

for all subsystems, and all $λ_i$ converge to the same optimal value. Thus, the local Lagrange Multiplier $λ_i$ can be computed as follows.

$$λ_i = - \nabla u_i J_{X,i} (\nabla u_i, g_i(x_i, u_i, d_i))^{-1}$$ (12)

where the number of local manipulated variables must be equal to or more than the number of constraints in common, and the solution must be unique.

In order to evaluate (12), each subsystem $i$ is required to estimate its local steady-state cost and constraint gradient, which can be achieved locally using any model-based or model-free gradient estimation. This estimation takes into account the effect of the updated input calculated in (11).
For a list of gradient estimation techniques for RTO see Srinivasan et al. (2011), and François et al. (2012).

Virtual subsystem, \( \lambda_0 \) is always 0 because \( J_{N,0} \) is defined as 0, and to limit the dual variable to be non-negative in a steady-state condition (i.e., to satisfy steady-state dual feasibility).

In a practical context, if a local problem is simple enough that it takes almost no time to find the local solution and no possible numerical issues, then there is no reason to avoid an equation solver to estimate the local Lagrange multiplier.

### 3.3 Online Primal Decomposition Framework

By combining the concept of primal decomposition, the idea of central constraint controllers, local setpoint controllers, and local Lagrange multiplier estimation as described above, we propose to solve the problem of real-time resource allocation in handling coupling constraint using distributed feedback-optimizing control using Primal decomposition framework. This framework theoretically can reach steady-state optimal condition and guarantees primal feasibility.

Fig. 1 illustrates the implementation of this framework in solving the above problem. The central constraint controllers, containing virtual, normal and compensator subsystems, provide new set points for local coupling constraint, \( g_i \) (see eq. (7), (9), and (10)). These set points will be tracked by local setpoint controllers (see eq. (11)). Should there be any disturbance \( d_i \), one can use the current plant information to estimate the plant’s current state and parameters/disturbance using local dynamic estimator such as Extended Kalman Filter (EKF). Using the inputs, estimated states and parameters/disturbance, one can estimate both cost and constraint gradient to compute the local Lagrange multipliers as shown in eq. (12). Thereafter, these multipliers are used to determine the new setpoints central constraint controllers.

![Fig. 1. The proposed online primal decomposition control structure for optimal operation using simple PID controllers and selector. The grey and white boxes represent the physical system and the computation block, respectively.](image)

**The generality:** Note that we have two types of shared resources here, the shared input, that we need to optimally allocate, and the shared constraint, \( g \), that all subsystem should cooperate optimally to satisfy. The shared input can be total flow of materials, steam, or energy that any kind of process industries usually need. The shared constraint can be any type of constraint that two or more subsystems have influence on, i.e., plant capacity. Specifically for the solid mining industry, we could consider the extracted earth deposit as the flow of materials, and the maximum capacity of the processing plant, e.g., smelter, as the shared constraint.

### 4. SIMULATION EXAMPLE

In this section, we apply the proposed method control structure on a gas-lifted well network (liquid and gas extraction activities) with \( N = 2 \) wells, that are operated locally. The optimization objective of this case is to maximize total oil production, \( w_{to} = \sum_{i=1}^{N} w_{po,i} \) while minimizing the cost of total gas lift, \( w_{gl} = \sum_{i=1}^{N} w_{gl,i} \). Thus, \( J_{N'} = -p_o w_{po,i} + p_g w_{gl,i} \), where \( p_o \) and \( p_g \) are the oil price and the gas lift cost, respectively. The coupling constraint is \( g(x, u, d) = \sum_{i=1}^{N} w_{pg,i} - N w_{po,i} \), where \( w_{pg,i} \) is the local produced gas, and \( N w_{po,i} \) is the maximum capacity to handle total produced gas. Fig. 2 illustrates this case study completed with the proposed control structure, where well 2 is assigned to ensure the setpoint primal feasibility (see Eq. (7)). Note that each well has an MPFM to measure the actual local produced gas.

![Fig. 2. A simplified process diagram of a gas-lifted oil production network with constraints in maximum produced gas handling capacity, and equipped with the proposed control structure. Dashed lines represent data transmission in the proposed control structure. Act stands for actuator. \( X \) represents controllers.](image)
also lead to an unconstrained case, where the coupling constraint (3c) is inactive.

The controllers are tuned using SIMC (Simple/Skogestad Internal Model Control) rules introduced by Skogestad (2003). The desired timescale of the local setpoint controllers is $\tau_{cL,1} = \tau_{cL,2} = 75$ seconds. To satisfy the timescale separation concept, the chosen timescale for the central constraint controllers is $\tau_{c1} = \tau_{c2} = 131.25$ seconds. Thus, the timescale ratio $\epsilon = \tau_{cL,1}/\tau_{c1} = 0.5714$. Since the time delay is insignificant, one could consider integrating controllers. However, we use PI controllers in this simulation. Tab 1 displays the controllers’ parameters.

| Gain $K_{f,0}$ | Time Constant [sec] $\tau_{f,0}$ | Time Delay [sec] $\theta_0$ | $K_{f,1}$ | $\tau_{f,1}$ | $\theta_1$ | $K_{fL,1}$ | $\tau_{fL,1}$ | $\theta_{1,1}$ | $K_{fL,2}$ | $\tau_{fL,2}$ | $\theta_{1,2}$ |
|----------------|---------------------------------|----------------------------|-----------|-------------|-------------|-----------|-------------|-------------|-----------|-------------|-------------|
| 0.375          | 617                             | 0                         | 0.375     | 617         | 0           | 1.031     | 643         | 9           | 1.039     | 614         | 2           |

To estimate the local Lagrange multiplier, we execute three steps. First, we use the current plant information to estimate the plant’s current state and parameters using EKF. Next, we use the updated model to evaluate the steady-state gradients. These first two steps utilize the same methods we use in Dirza et al. (2021). Finally, we evaluate Eq. (12).

First, we solve the integrated production optimization problem (3) to obtain the ideal steady-state optimal setpoint as the baseline. Then, we implement the proposed framework described in Section 3.

Fig. 3 shows the simulation results of the produced gas setpoints. These are the output of the central constraint controllers’ performance, where we can observe that the total setpoint of the produced gas is not violating the constraint. As a consequence, the compensator subsystem (subsystem 2) ‘absorbs’ the violation, indicated by oscillations during transient. These associated oscillations can also be observed in Fig. 4-5. Moreover, the produced gas setpoint of each well reaches the steady-state optimal setpoint labeled by $w_{pg,i}^{sp}$. Furthermore, the steady-state slack variable $w_{pg,0}^{sp}$ also reaches 0 in the constrained case and $w_{pg,0}^{sp} > 0$ in unconstrained case.

Fig. 4 depicts the simulation results showing the performance of the local setpoint controllers and the local Lagrange multiplier estimator. The top plot shows that the local setpoint controllers have successfully tracked the produced gas setpoints given by the central constraint controllers. The middle one shows that the manipulated variable, i.e., gas-lift rates, reached the optimal steady-state conditions. Additionally, the local Lagrange multipliers converge to the optimal steady-state conditions and satisfy the dual feasibility in the steady-state shown in the bottom plot. Note that, in unconstrained case, the steady-state local Lagrange multiplier is 0. These results confirm the applicability of using a virtual variable to store the unutilized produced gas handling capacity.

Fig. 5 displays the simulation results for the actual cost and the produced gas. The top plot shows that the presented method can reach the optimal steady-state cost. Moreover, the total produced gas satisfies the constraint with relatively short duration and insignificant magnitude violations during the transients. As mentioned above, this violation only depends on the tuning parameter we choose in the local setpoint controllers because the central constraint controllers have given the setpoints that guarantee the primal feasibility (see Fig. 3).

Fig. 6 shows the comparison with dual decomposition used in Dirza et al. (2021), where the central constraint controllers has to be slower (in timescale) than the presented method in active constraint switching (i.e., unconstrained to a constrained case). This requirement may lead to dynamic constraint violation when local gradient controllers are too aggressive, whereas central constraint controllers of the dual approach has no specific strategy to regulate the primal feasibility. During the transient, the Lagrange multiplier is suboptimal. This condition significantly con-
point controllers, which can be tuned more independently.

violation is purely the product of the aggressive local set-point to satisfy the constraint. Thus, the 'small' constraint controller of the presented approach ensures the attributes to constraint violation. Unlike dual, the central

Fig. 6. Comparison with different timescales separation (\(\epsilon\)) and method. Small \(\epsilon\) means the method of interest has slow timescale central constraint controller, and vice versa.

tributes to constraint violation. Unlike dual, the central constraint controller of the presented approach ensures the total setpoint to satisfy the constraint. Thus, the 'small' violation is purely the product of the aggressive local setpoint controllers, which can be tuned more independently. Even this method does not need local setpoint controllers when it only has input constraints. Meanwhile, the dual approach may have an issue in tight constraint control as the central constraint controller has to be in a slow timescale. Forcing a faster timescale central constraint controller (larger \(\epsilon\)) may lead to oscillatory behavior.

When it comes to solid extraction activities, one may consider a network of mines and smelters in a metal mining industry (Wei et al., 2004), which fundamentally has similar class of problem as the gas-lift well network. These mines produce concentrates that should be sold and transported to the smelters for processing. The optimization problem is to optimally allocate the raw materials from the mines in order to achieve maximum revenue since production capacity of the smelters is limited.

5. CONCLUSION

In this paper, we presented a real-time optimal resource allocation using the framework of online primal decomposition. We showed that such a problem turns into a feedback control problem by introducing virtual subsystems or slack variables to store unutilized resources, implementing central constraint controls and local setpoint controls, and estimating Lagrange multipliers. The goals of central constraint controls are to directly control the constraint, update the local constrained variables setpoints, and regulate the primal feasibility of the constrained variables. The objective of local setpoint controls is to control constrained variables to the given setpoint. For the case study we consider in the simulation example, this proposed framework leads to a system-wide optimal operation without a numerical solver. Moreover, the setpoints provided by the central constraint controls satisfy the primal feasibility. Furthermore, similar class of problems from different industries, i.e., metal, mining, and metal processing, can also be solved using the proposed framework.

REFERENCES

Baldea, M. and Daoutidis, P. (2007). Control of integrated process networks—A multi-time scale perspective. Computers & Chemical Engineering, 31(5), 426–444.

Boyd, S., Xiao, L., Mutapcic, A., and Mattingley, J. (2008). Notes on Decomposition methods.

Dirza, R., Skogestad, S., and Krishnamoorthy, D. (2021). Optimal Resource Allocation using Distributed Feedback-based Real-time Optimization. IFAC-PapersOnLine, 54(3), 706–711.

Engell, S. (2007). Feedback control for optimal process operation. Journal of Process Control, 17(3), 203–219.

François, G., Srinivasan, B., and Bonvin, D. (2012). Comparison of six implicit real-time optimization schemes. Journal européen des systèmes automatisés, 46, 291–305.

Hansen, L.S., Pedersen, S., and Durdevic, P. (2019). Multi-Phase Flow Metering in Offshore Oil and Gas Transportation Pipelines: Trends and Perspectives. Sensors, 19(9).

Jäschke, J., Cao, Y., and Kariwala, V. (2017). Self-optimizing control – A survey. Annual Reviews in Control, 43, 199–223.

Krishnamoorthy, D. (2021). A distributed feedback-based online process optimization framework for optimal resource sharing. Journal of Process Control, 97, 72–83.

Maestre, J.M., Negenborn, R.R., and others (2014). Distributed model predictive control made easy, volume 69.

Morari, M., Arkun, Y., and Stephanopoulos, G. (1980). Studies in the synthesis of control structures for chemical processes: Part I: Formulation of the problem. Process decomposition and the classification of the control tasks. Analysis of the optimizing control structures. AIChe Journal, 26(2), 220–232.

Shook, D. (2006). Best practices improve control system performance. Oil and Gas Journal, 104(38), 52–54.

Skogestad, S. (2003). Simple analytic rules for model reduction and PID controller tuning. Journal of Process Control, 13(4), 291–309.

Srinivasan, B., François, G., and Bonvin, D. (2011). Comparison of Gradient Estimation Methods for Real-time Optimization. In 21st ESCAPE, volume 29 of Computer Aided Chemical Engineering, 607–611. Elsevier.

Wei, Y.M., Hu, Q.h., and Fan, Y. (2004). Mathematical model for the optimization of the allocation of nonferrous raw materials in China. Computers & Industrial Engineering, 46(2), 293–303.

Wenzel, S., Paulen, R., Stojanovski, G., Krämer, S., Beisheim, B., and Engell, S. (2016). Optimal resource allocation in industrial complexes by distributed optimization and dynamic pricing. at - Automatisierungstechnik, 64(6), 428–442.