Draft for ESCAPE Paper

WHAT I'M TRYING TO CONVEY TO THE READER

ACTUAL TEXT IN PAPER - (DRAFT)

UNRESOLVED ISSUES/DOUBTS IN MY HEAD THAT MAKE ME UNEASY.

Multi-scenario Design Optimization using ADMM of a Thermal Energy Storage system

(0.25 page) **Abstract** →

TBC

(0.5 pg) **SEC 1** → INTRODUCTION.

TBC

(2 pg) **SEC 2** – METHODOLOGY / DECOMPOSING TWO-STAGE DYNAMIC OPTIMIZATION PROBLEMS USING ADMM

Explain how 2-stage D.O problems can be decomposed as needed in our approach.

> The optimal design problem for a dynamic system under time varying uncertainty can be represented as a 2-stage dynamic optimization problem.

Explain the general notation for the central problem

The design decisions being the first stage variables (w^{des}) the operations decisions as the second stage variables (w^{oper})

In this representation it is assumed that the uncertainty is fully realized after the first stage decisions (design decisions) are made and hence the operations are optimal for the realized uncertainty.

We can represent the uncertainty by a set of discrete scenarios $S = \{1, 2, \dots S\}$ with the cost weight ω_s to represent the likelihood of scenario s being realized.

The dynamics of the system can be discretized into N equally spaced sampling intervals represented by the set $K = \{0, 1, \dots N-1\}$.

The optimal design problem can be then cast as a nonlinear programming problem (NLP) of the form

$$\min_{w^{des}, w^{oper}} \quad \phi^{capex}(w^{des}) + \sum_{s \in S} \omega_s \left[\sum_{k \in K} \ell(x_{k,s}, \quad u_{k,s}) \right] \qquad (1)$$

$$\text{s.t.} \quad g(x_{k,s}, \quad u_{k,s}, \theta_{k,s}, w^{des}) \leq 0 \qquad \forall k \in K, s \in S$$

$$x_{k+1,s} = h(x_{k,s}, u_{k,s}, \theta_{k,s}, w^{des}) \qquad \forall k \in K, s \in S$$

$$x_{0,s} = x_0 \qquad \forall s \in S$$

$$(w^{oper})^T := [x_{0,s}^T, x_{k+1,s}^T, \quad u_{k,s}^T] \qquad \forall k \in K, s \in S$$

→ How to make stacking more easier to read?

where the $(\cdot)_{k,s}$ represents the s^{th} scenario at time step k.

Vectors x and u represent the differential states and the control inputs applied to the plant during operations respectively.

The initial condition during operations is given by x_0 for all the scenarios

Explain variables. →

In the interest of simplifying notation we collect these second stage variables for all the scenarios together into a single vector (w^{oper})

The time varying parameters are represented by the vector θ.

Explain functions →

The function $\phi^{capex}(\cdot)$ is used to represent the capital cost and $\ell(\cdot)$ represent the operating cost.

The objective of the design problem is then simply defined as the sum of capital costs and the expected costs during operation.

Function $g(\cdot)$ represent the nonlinear inequality constraints and $h(\cdot)$ used to represent the dynamics of the system.

Sec 2.1 → Reformulating the design problem as a general form consensus optimization problem.

Explain how the general decomposed problem will look like.

Solving the optimal design problem as formed in (1) can become computationally intractable due to memory constraints when having to consider many scenarios for uncertainty and long horizon times.

smaller partition problems which can then be solved seperately a way to handle the memory limitations.

We can divide the two stage dynamic optimization problem into P partitions denoted by the set $\mathbb{P} = \{1, 2, \dots P\}$ in a very flexible manner.

An illustrative example of a design problem with 2 scenarios divided into 6 partition problems is shown in Figure 1.

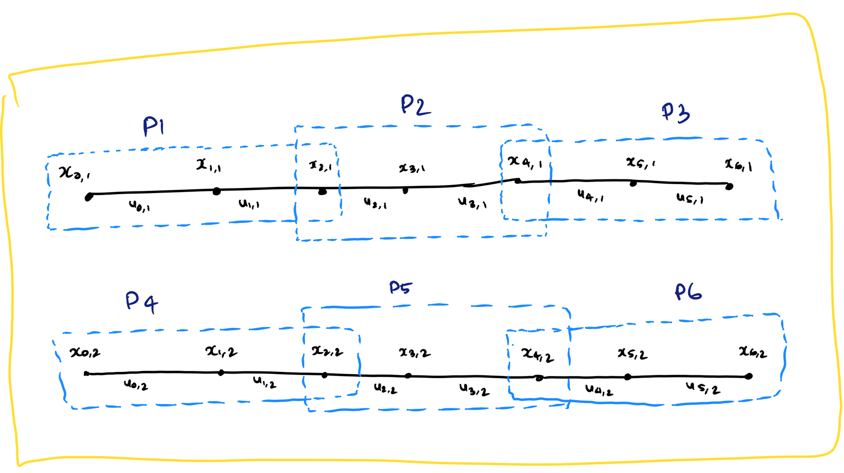


Fig 1 → An illustration of dividing the optimal design problem with 2 scenarios into 6 partitions

In the partition problems, each partition p has its own local variables for design (w_p^{des}) and operations (w_p^{oper})

The sum of the individual partition problems together is similar to the central problem in (1) when, } Still feels strange.

1) All the partitions are in consensus of the value of their local design variables,

2) Adjacent partitions are in consensus of the differential state variable shared between them (at the edge of the partitions)

We can introduce a global copy of all the variables that must reach consensus into a vector denoted as v.

The consensus requirements can then be imposed as an additional constraint in each partition p as

$$w_p^{des} - v^{des} = 0$$
$$A_p w_p^{oper} - B_p v^{oper} = 0.$$

where A_p and B_p are selection matrices used to link a subset of the local variables of partition p to the corresponding sections in the global copy (v)

We can thus write the optimal design problem as a sum of smaller partition problems that are linked with consensus constraints as,

$$\min_{w_p, v} \quad \sum_{p \in \mathbb{P}} \Phi_p^{CAPEX}(w_p^{des}) + \Phi_p^{oper}(w_p^{oper}) \qquad (2)_a$$

$$\text{st.} \quad g_p(w_p^{oper}, \theta_p, w_p^{des}) \leq 0 \qquad \forall p \in \mathbb{P} \qquad (2b)$$

$$h_p(w_p^{oper}, \theta_p, w_p^{des}) = 0 \qquad \forall p \in \mathbb{P} \qquad (2c)$$

→ Cannot do the x^{k+1} form without creating a mess

$$w_p^{des} - v^{des} = 0 \qquad \forall p \in \mathbb{P} \qquad (2d)$$

$$A_p w_p^{oper} - B_p v^{oper} = 0 \qquad \forall p \in \mathbb{P} \qquad (2e)$$

$$w_p \doteq [w_p^{des}, w_p^{oper}] \qquad \forall p \in \mathbb{P} \qquad (2f)$$

The objective function terms in the partition problems are chosen appropriately to add up to the original objective in (1)

For the sake of compact notation, we collect the local variables in the partition as w_p and can represent the constraints (2b) and (2c) which only involve the local variables (w_p)

SEC 2.2 → APPLYING ADMM to get a Distributed algorithm.

The individual partition problems are not trivially seperable due to the presence of constraint

of (2) can be formed by taking the the constraint (2d) and (2e) to the objective as

$$\mathcal{L}_p(w_p, v, \lambda_p) = \sum_{p \in \mathbb{P}} \left[\Phi^{capex}(w_p^{des}) + \Phi_p^{oper}(w_p^{oper}) + (\lambda_p^{des})^T (w_p^{des} - v^{des}) + \frac{\rho^{des}}{2} \| w_p^{des} - v^{des} \|_2^2 + (\lambda_p^{oper})^T (A_p w_p^{oper} - B_p w_p^{oper}) + \frac{\rho^{oper}}{2} \| A_p w_p^{oper} - B_p v^{oper} \|_2^2 \right] \qquad (3)$$

where $\lambda_p^T := [(\lambda_p^{des})^T, (\lambda_p^{oper})^T]$ are the lagrange multipliers associated with the consensus constraints ((2d) and (2e)) and $\rho^T = [(\rho^{des})^T, (\rho^{oper})^T]$ is a vector of chosen penalty parameters.

It can be seen that (3) is additively seperable except for the quadratic penalty terms.

The i^{th} iteration of ADMM algorithm involves solving the partition problems while keeping v and λ fixed and then updating them by keeping the local variables w_p fixed in an alternating fashion untill convergence

The iteration i of the ADMM algorithm thus takes the form

$$(w_p)^{i+1} = \underset{w_p \in \mathbb{W}_p}{\operatorname{argmin}} \; \mathcal{L}_p(w_p, v^i, \lambda_p^i) \qquad \forall p \in \mathbb{P} \qquad (4a)$$

$$(v)^{i+1} = \underset{v}{\operatorname{argmin}} \; \mathcal{L}_p(w_p^{i+1}, v, \lambda_p^i) \qquad (4b)$$

$$(\lambda_p^{des})^{i+1} = (\lambda_p^{des})^i + \rho^{des}\left((w_p^{des})^{i+1} - (v^{des})^{i+1}\right) \qquad \forall p \in \mathbb{P} \qquad (4c)$$

$$(\lambda_p^{oper})^{i+1} = (\lambda_p^{oper})^i + \rho^{oper}\left(A_p (w_p^{oper})^{i+1} - B_p (v^{oper})^{i+1}\right) \qquad \forall p \in \mathbb{P} \qquad (4d)$$

We can see that steps (4a) involve solving an optimization problem for each partition seperately while the global variable is kept constant.

Thus partition problem can be solved using its own seperate machine

Step (4b) is the minimization of the AL function while the local variables in the partitions are kept constant.

for consensus optimization problems reduces to finding the minimum of a quadratic function and can be shown to be the averaging operator

$$v^{i+1}(j) = \frac{1}{|\mathbb{P}_j|} \sum_{p \in \mathbb{P}} w_p^{i+1}(j)$$

where $\mathbb{P}_j \subseteq \mathbb{P}$ denotes the set of partitions connected to the global $v(j)$

Steps (4c) and (4d) are the updates to the lagrange multipliers of the consensus constraints and can also be carried out in each partition seperately.

Explain termination criteria for ADMM →

termination criteria used for the ADMM iterations are that the primal and dual residuals are reasonably close to zero as explained in (Boyd)

$$\| r^{i+1} \|_2 \leq \epsilon^{primal}$$
$$\| s^{i+1} \|_2 \leq \epsilon$$

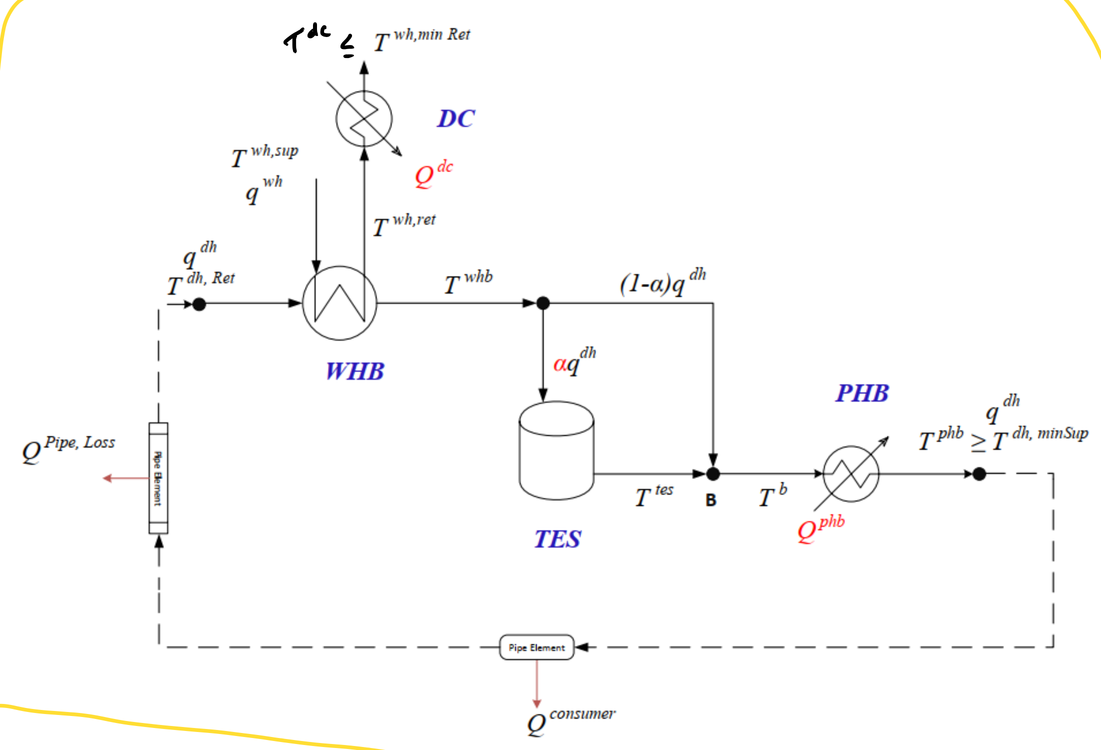
The primal and dual residuals in our case are defined as

$$r_p^{i+1} = \begin{bmatrix} (w_p^{des})^{i+1} - (v^{des})^{i+1} \\ A_p (w_p^{oper})^{i+1} - B_p v^{oper} \end{bmatrix} \qquad \forall p \in \mathbb{P}$$

$$s_p^{i+1} \quad (\lambda_p^{des})_{i+1} - (\lambda_p^{des})_i \qquad \forall p \in \mathbb{P}$$

(1 pg) SEC 3 → The TES design Problem.

Describe the Process we have. → we use the flowsheet in Fig 2 to represent the heating section of a district heating network where water is used in a closed loop to satisfy the heating requirements of households in an area.



→ Do i need to change? Details are there, does not look too professional academic like dinesh's ones - since made in Visio

Fig 2 ⇒ Simple flowsheet of a district heating system

The volumetric flowrate of water in the loop (q^{dh}) is determined by the number of consumers and their heating demands at any particular time.

The temperature of the water returned to the heating section ($T^{dh,Ret}$) is determined by the heat losses in the system and is assumed to be correlated to the ambient weather conditions. → Need to explain more/better about the suppliers?

The operational objective of the heating system is to heat this water up to a temperature (T^{Phb}) which must be above a contractually specified temperature ($T^{dh,minRet}$).

A cheap source of heat is available in the form of a process stream from an industrial process which requires to be cooled down below a specified temperature ($T^{wh,min\ Ret}$).

This stream can loose heat to the district heating system using the heat exchanger WHB, and any additional cooling required in the process stream satisfied by the utility heat exchanger DC which uses some external cooling utility.

In a similar manner, any additional heating required in the district heating side is satisfied by using the peak heat boiler (PHB) which burns fossil fuels.

There is a temporal mismatch in the supply of heat from the process stream and the demand for heating from the consumers.

In order to better manage this mismatch and decrease the reliance on external utilities, a thermal energy storage system in the form of a simple buffer tank is being considered to be installed.

This simple TES system can charge/discharge thermal energy by raising/lowering the temperature of the tank by manipulating the flow split α.

Explain model simplifications used → The TES tank is modelled as a well mixed tank with perfect level control with the volume maintained at V^{tes}

The WHB is modelled as a purely countercurrent heat exchanger with the hot and cold sides discretized into a series of N_{cell} cells. Each cell is assumed well mixed with the heat transfer occurring only across the cell walls.

Heat exchangers PHB and DC are modelled simply as a well mixed tank with the heat duty added or removed directly

How the equations are obtained → Mass and energy balance equations can thus be written out for the system to model the dynamics of the system.
The differential states thus consist of $x^T = [T^{whb}, T^{wh,min\ Ret}, T^{tes}, T^{phb}]$
while the control inputs are $u^T = [\alpha, Q^{phb}, Q^{dc}]$

The scenarios → what are the θ_k's? In our model, we consider the uncertain parameters at each time step k as $\theta_k = [q^{dh}_k, T^{dh,Ret}_k, q^{wh}, T^{wh,sup}]$
The uncertainities of the future profile can then be represented by building representative scenarios of these uncertain parameters.

what is the aim of the design problem. → The design problem is then to find the optimal volume of the TES tank (V^{tes}) and the area of heat exchanger (A^{whb}) that must be installed, given the uncertanities in future supply and demand of thermal energy.

We can thus form it as the large two-stage dynamic optimization problem of the form in eqn (1), partition it into smaller problems of the form shown in eqn (2), which can then be solved in a distributed manner using ADMM as described in Sec 2.2

(1 pg) SEC-4 → RESULTS AD DISCUSSIONS

We present a simple simulation study to demonstrate the approach described in section 2 on the TES model described in Section 3 and compare the solution and some convergence metrics against solving the design problem as a single NLP (i)

Exact numbers used in sim not given since, → No real basis used for the numbers to make my life easier → A lot of params and not enough space.

We consider a simple case with 2 equally probable scenarios with $N = 60$ discretizations to represent the future operations.

The profiles of q^{wh} and $T^{wh,sup}$ both have a step change at time steps $k = 20$ and $k = 40$, while all the other parameters are held at their steady state values

The design variables (w^{des}) and operations variables (w^{oper}) were all scaled to be within zero and one while formulating the optimization problem.

Quadratic functions are used to represent the capital cost $\left(\phi^{capex} = 10^{-3} (V^{tes})^2 + 10^{-3} (A^{whb})^2 \right)$ and the operating costs $\left(\ell = (q^{phb})^2 + (q^{dc})^2 \right)$

Have an urge to make into more realistic units and get basis. But → No time to do → No space to explain. Hopefully reviewers not too critical in this

Scaling facilitated selecting the AL penalty parameter ρ to tune the convergence speed of our distributed approach more easily.

The penalty parameters chosen as $\rho^{des} = 1e^{-3}$ and $\rho^{oper} = 1e^{-1}$ in the partition problems to roughly balance the magnitude of the objective and penalization terms.

We solve the problem using the distributed algorithm in sec 2.2 and show the results against the optimal design solution by solving the problem centrally.

Exact volume and Area to be confirmed. with the plots.

Want the reader to see that the residuals are going to zero quickly and how the same optimal design is reached

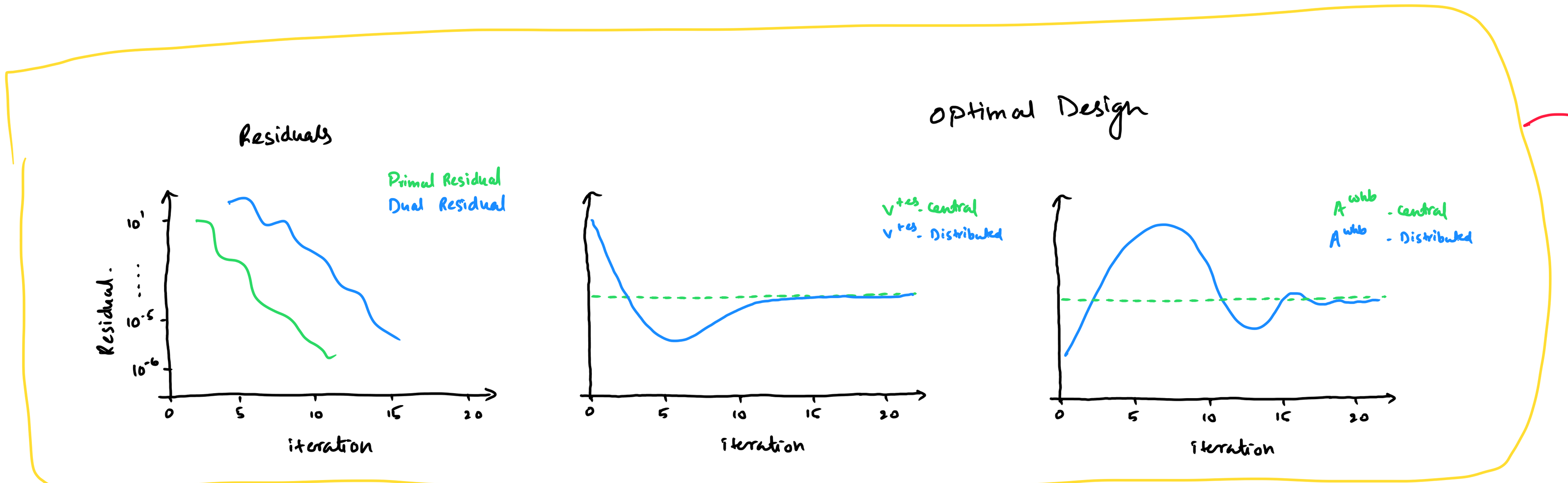


Fig 3 → Results from solving the problem centrally vs using the distributed approach using ADMM.

From Fig 3a, we can see that the primal and dual residuals decrease with ADMM iterations indicating convergence of the ADMM iterations.

From Fig 3b&c, The optimal design variables can be seen to converge to the same solution as the one found by solving the problem centrally.

Disclaimer about ADMM in case of non convex problems. →

An important point to note is that this behaviour is not guarenteed by ADMM in the case of nonconvex problems.

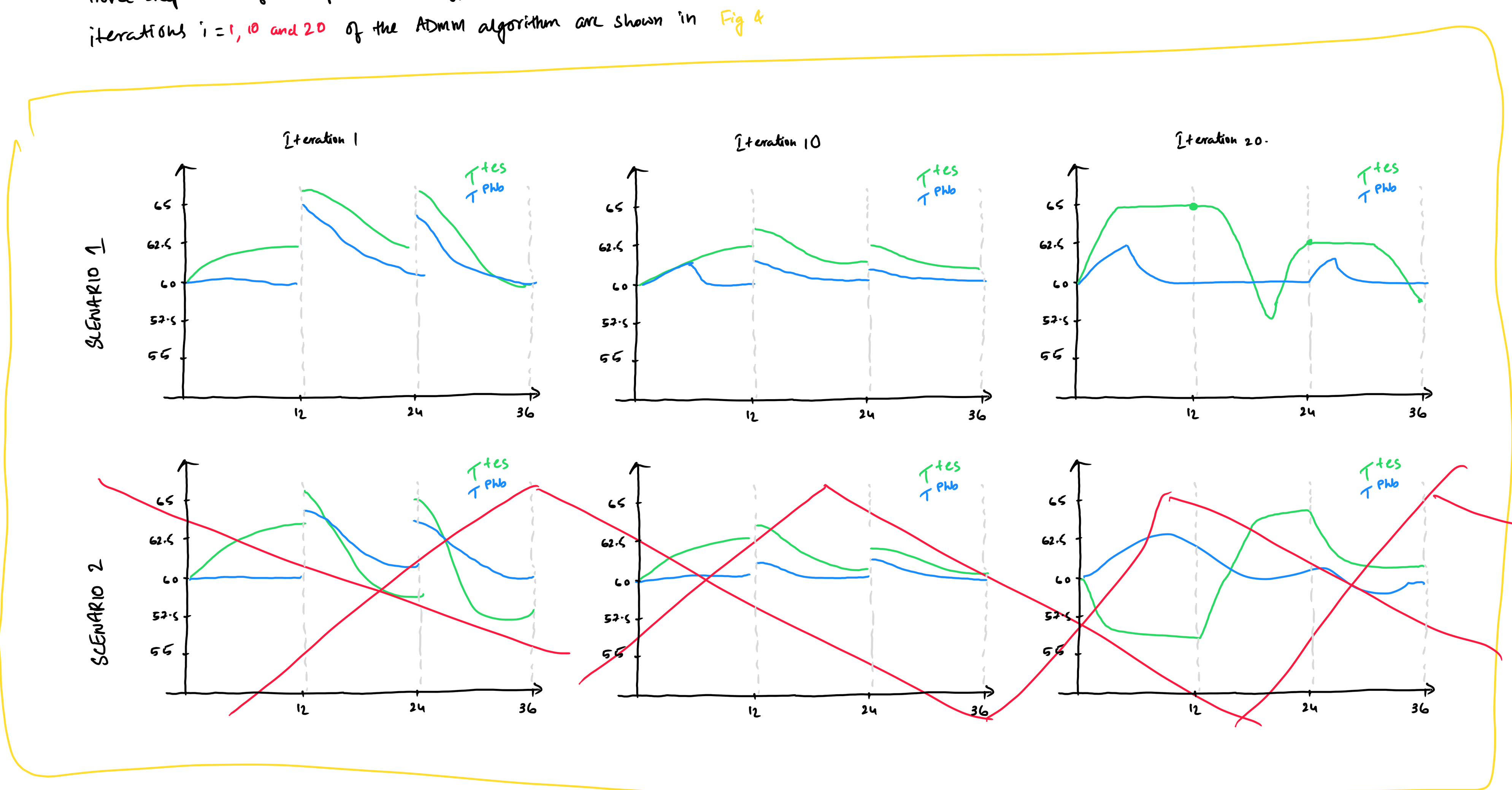
When applied to nonconvex problems, ADMM need not converge, and even when it does converge, it need not converge to an optimal point and must be hence considered just as another local optimization method (Boyd 2011)

Although the convergence guarantees for ADMM in the case of complex nonconvex NLPs are poorly understood, it has been shown to perform satisfactorily in practice (Rodriguez)

Three snapshots of the optimal differential state variables in the partition problems at iterations $i = 1, 10$ and 20 of the ADMM algorithm are shown in Fig 4

Step Profiles to be confirmed.

Want the reader to see how stitching happens in the differential states as ADMM iterates



Deleting to save paper real estate.

Fig 4 → Achieving consensus in the second stage differential states between the partition problems

Discussion on similarities with MS →

An interesting observation in our approach is that, partition 2 and 3 have the initial condition of the dynamic optimization problem as a local variable.

In Figure 4, we can see that during initial iterations (left subplot) the optimal solution is to start the TES tank at the maximum allowable temperature.

The penalization terms added are then able to implicitly close this gap between the partitions and reach consensus. (middle and right subplot)

In this aspect our approach shares similarities to multiple shooting (MS) approach in dynamic optimization.

The key difference is that the state continuity (consensus constraint) is enforced explicitly by the solver in MS, while in the ADMM approach it is enforced implicitly by minimizing the AL.

Discussion about speed → The distributed approach converges in 20 iterations of the ADMM algorithm, where smaller NLPs are solved in each iteration.

When comparing against the centralized solution, we do observe that our distributed approach takes more computation time, even if perfect parallelization is assumed.

→ Thinking of removing this whole section bcz → Real estate in paper → Speed & CPU etc a pandoras box i do not want to open now.

→ Convergence takes many iterations (Superlinear)
→ An easy and intuitive approach without the need for specialized solvers or, with minimal changes to existing models and workflows.
→ Speed can be improved by decomposing at the linear algebra level, but need specialized implementation of the solvers (Cont latest technology)

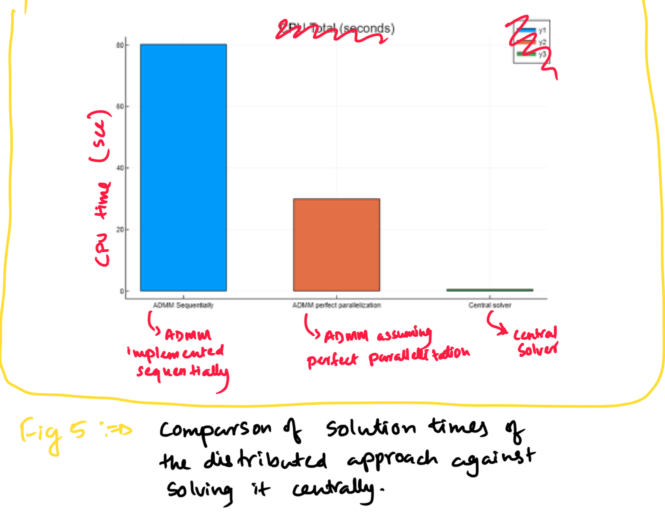


Fig 5 → Comparison of solution times of the distributed approach against solving it centrally.

This is due to the fact that the optimizer does not have access to the full information of all the other partition problems and hence cannot take advantage of it.

Using internal decomposition techniques would be more suited if solution speed is of interest along with memory limitations.

The use of internal decomposition methods for forming general partitions for similar two-stage dynamic optimization problems would be an interesting area for future work.

(0.25 pg) SEC-5 - CONCLUSIONS

In this paper, we presented the optimal design of a simple TES system under uncertainity as a two-stage nonlinear dynamic optimization problem.

Due to limitations in memory of solving the problem centrally in a single machine, an approach for forming smaller partition problems in a general fashion was shown.

The ADMM algorithm was applied to coordinate between the subproblems which could be solved seperately and in parallel.

A simple simulation exercise was used to demonstrate the approach.

REFERENCES

→ Boyd → Statistical learning and ADMM
→ Rodriguez → ADMM in nonconvex NLPs
→ Mandar → (Escape) - optimizing capacity of TC