On Coordinated Primal-Dual Interior-Point Methods for Multi-Agent Optimization

Altuğ Bitlislioğlu and Colin N. Jones

Abstract— **This paper presents a coordinated primal-dual interior point (PDIP) method for solving structured convex linear and quadratic programs (LP-QP) in a distributed manner. The considered class of problems represents a multi-agent setting, where the aggregated cost is to be minimized while respecting coupling constraints as well as local constraints of the agents. Unlike fully distributed methods, a central agent is utilized, which coordinates the Newton steps taken within the interior-point algorithm. In practical PDIP implementations, predictor-corrector variants are widely used, due to their very fast convergence. We show that in the coordinated case, a naive implementation of a PDIP with predictor-corrector scheme introduces extra communication steps between local and central agents. We propose a decentralized predictor-corrector scheme that uses a non-uniform perturbation on the complementary slackness condition, which is able to reduce the number of communication steps while preserving fast convergence of the original methods. The proposed coordinated PDIP method with decentralized predictor-corrector scheme can be analysed in the general framework of PDIP methods with non-uniform complementarity perturbations, for which convergence and complexity results are provided.**

I. Introduction

We consider the following convex multi-agent optimization problem:

$$\begin{aligned}
\underset{x^1,\dots,x^m}{\min.} \quad & \sum_{i=1}^{m} f^i(x^i) \\
\text{s. t.} \quad & \forall i \in [1,\dots,m] \ , \\
& g^i(x^i) \leq 0 \ , \\
& C^i x^i = y^i \ , \\
& \sum_{i=1}^{m} D^i y^i = d \ ,
\end{aligned} \tag{1}$$

where superscript i indicates that the function or variable belongs to agent i. The cost function and the constraint set are separable, except for the last equality constraint which introduces coupling between agents. We write (1) in the 'output coupling' format, in order the distinguish the local variables x^i, which determine feasibility and the cost of the agent, and the output variables y^i, which contribute to the coupling between agents.

Problem (1) captures a wide area of interest in estimation and control in networks [17]. Distributed control problems with dynamic or output coupling between agents can be formulated in the form of (1). Coupling in the cost and inequality constraints can also be incorporated into (1) by reformulations with auxiliary variables. See [17] for a survey on distributed problems on estimation and control and categorization of formulations and solution methods.

Distributed and coordinated methods for solving problems of the form (1) is an active area of research. We distinguish here 'coordinated' schemes from 'distributed' schemes by the existence of a central agent that can collect information from all agents, inducing a 'star' topology on the information network. Many engineering applications, such as power grids and chemical plants, demand a central point of information, e.g. SCADA systems, for enabling fault surveillance, intervention by human operators and data storage. Furthermore, without a central agent, the responsibility of modelling and managing the coupling between agents will fall on the agents, who might not have the domain knowledge. Another motivation for having a central agent is that the convergence of the optimization method will typically be much faster compared to fully distributed schemes since information can be shared efficiently. Distributed methods can be applied as a coordinated algorithm by choosing the splitting of the problem accordingly. On the other hand coordinated methods can often be made fully distributed if it is possible to allocate the task of the central node among all agents.

Coordinated solution methods are strongly linked to decomposition methods where the problem is split into a master problem and many sub-problems. Dual decomposition methods use Lagrangian relaxation and do not provide feasibility until convergence, whereas primal decomposition methods decompose the problem directly and preserve feasibility. In the constrained setting, the master problem becomes non-differentiable and is usually solved with suitable first order methods such as sub-gradient or cutting plane methods [3]. Applying second order methods is still possible by smoothing the problem with barrier functions, either in the dual setting [13], [18], [25] or in the primal setting [5], [16], [24]. Using second order curvature information, these interior-point decomposition methods become robust against scaling and can converge faster to high accuracy solutions compared to first order methods [5], [17].

Apart from primal and dual interior-point decomposition methods, there is a third option: applying a primal-dual interior point method (PDIP) [23] directly to the problem and decomposing the Newton step. This primal-dual scheme was proposed in [1], [20], where authors focus on distributed solutions of the Newton step to eliminate the need for a

This work has received support from the European Research Council under the European Unions Seventh Framework Programme (FP/2007-2013), ERC Grant Agreement 307608.

The authors are with Laboratoire d'Automatique, École Polytechnique Fédérale de Lausanne (EPFL), Switzerland .
`altug.bitlislioglu - colin.jones @ epfl.ch`

central agent, using proximal splitting [1] and message passing [20] methods. The advantage of the primal-dual interior-point setting is two fold: Firstly there is no distinction between local and global iterates and iterations for solving local problems which only generate intermediate Hessian and gradient information are eliminated. Secondly, by applying predictor-corrector schemes [10], [15], PDIP methods can become very efficient and therefore are commonly used in implementations aiming for high-performance [2], [9], [14]. However, in the interior-point decomposition framework, none of the aforementioned works apply a predictor-corrector scheme for enhancing convergence speed, with the exception of [13] which apply predictors in the dual decomposition setting, with primal-dual iterations in the local agents. Note that the method of [13] does not fall into the primal-dual framework as we use here.

For an algorithm used in the multi-agent setting, not only the number of iterations, but also the number of communications is of utmost importance for practical applicability. Even though coordinated algorithms can achieve the same number of iterations as a central method, the number of communication steps per iteration can be significant. We will show that applying Mehrotra's predictor-corrector scheme [15] in the PDIP setting, results in four communication steps per iteration. If the communication delay is significant, this might nullify the benefit of reduced number of iterations. To tackle this issue we introduce a decentralized predictor-corrector scheme in the coordinated PDIP setting, which enjoys the high performance of widely used predictor-corrector (PC) methods [10], [15], without introducing too many communication steps. We further analyze the algorithm from a general point of view of primal-dual interior point with non-uniform complementarity perturbations, or barrier parameters. This analysis provides safeguards that can be used to ensure convergence of the decentralized PC scheme.

II. Preliminaries

In the development of coordinated interior-point method, we will rely on the following assumptions regarding the problem data

Assumption 1: The local cost functions f^i's are convex.

Assumption 2: The local constraint sets $\{x^i : g^i(x^i) \leq 0\}$ are convex and compact.

Assumption 3: There exists a strictly feasible point for problem (1).

Assumption 4: The local output matrices, C^i's, have full row rank.

The first two assumptions ensure convexity of the optimization problem. The third assumption is necessary for using interior-point methods, whereas the last assumption will facilitate the decomposition of the problem with respect to the output variables.

A. Primal-Dual Interior Point Method

This section provides a summary of standard primal-dual interior point framework. See [19], [21] for a more in depth treatment and [11] for an up to date survey.

For brevity we re-write (1) in a compact form with slack variables.

$$\begin{aligned} \min_{x,y,s} \quad & f(x) \\ \text{s. t.} \quad & g(x) + s = 0 \ , \\ & \mathbf{C}x - y = 0 \ , \qquad (\mu) \\ & \mathbf{D}y = d \ , \qquad (\gamma) \\ & s \geq 0 \ , \qquad (\lambda) \end{aligned} \tag{2}$$

where the dual variables μ, γ, λ are shown next to their corresponding constraints and the following definitions are used:

$$s = \begin{bmatrix}(s^1)^T & \dots & (s^m)^T\end{bmatrix}^T \ ,$$

$$f(x) = \sum_{i=1}^{m} f^i(x^i), \quad g(x) = \begin{bmatrix}g^1(x^1)^T & \dots & g^m(x^m)^T\end{bmatrix}^T$$

$$\mathbf{C} = \text{blkdiag}(C^1, \dots, C^m), \quad \mathbf{D} = \begin{bmatrix}D^1 & \dots & D^m\end{bmatrix}$$

and y, x, d, μ, λ having similar definitions to s.

The first order optimality conditions for problem (2), or the Karush-Khun-Tucker(KKT) system, can be written as

$$r(x, y, s, \lambda, \mu, \gamma) = 0. \tag{3}$$

where $r = (r_{\text{dual}}^T, r_{\text{comp}}^T, r_{\text{prim}}^T)^T$, $r_{\text{dual}} = (r_{\text{dual},x}^T, r_{\text{dual},y}^T)^T$ and r_{prim} can be described similarly with the following definitions

$$\begin{aligned} r_{\text{dual},x}(x, \mu, \lambda) &= \nabla f(x) + \mathbf{C}^T\mu + \mathcal{J}_g(x)^T\lambda \ , \\ r_{\text{dual},y}(\gamma, \mu) &= \mathbf{D}^T\gamma - \mu \ , \\ r_{\text{prim,in},x}(x, s) &= g(x) + s \ , \\ r_{\text{prim,eq},x}(x, y) &= \mathbf{C}x - y \ , \\ r_{\text{prim,y}}(y) &= \mathbf{D}y - d \ , \\ r_{\text{comp}}(\lambda, s) &= \Lambda S\mathbf{1} \ . \end{aligned} \tag{4}$$

PDIP methods operate by applying Newton's method on a smoothed version of the nonlinear set of equations (3). The smoothing is obtained by perturbing the complementary slackness condition:

$$r_{\text{cent}}(\beta, \lambda, s) = \Lambda S\mathbf{1} - \beta, \quad \beta > 0 \ , \tag{5}$$

and using the modified function

$$\hat{r}(\beta) = (r_{\text{dual}}^T, r_{\text{cent}}^T(\beta), r_{\text{prim}}^T)^T .$$

In order to retain the optimizer of the original KKT system, the perturbation parameter is decreased progressively. As $\beta \to 0$, the primal-dual pairs converge to the optimizer of (2) following the so called 'central path'. In general the perturbation vector β is obtained by multiplying a scalar with a vector of ones. However we will let it attain different values for each constraint, similar to [8], which will allow us to decentralize the computation in Section III-B. The Newton direction $\Delta z = (\Delta x, \Delta y, \Delta s, \Delta\lambda, \Delta\mu, \Delta\gamma)$, for the perturbed KKT system can be found by solving the following linear system of equations:

$$\begin{bmatrix} H(x,\lambda) & 0 & 0 & \mathcal{J}_g(x)^T & \mathbf{C}^T & 0 \\ 0 & 0 & 0 & 0 & -\mathbf{I}_\lambda & \mathbf{D}^T \\ 0 & 0 & \Lambda & S & 0 & 0 \\ \mathcal{J}_g(x) & 0 & \mathbf{I}_s & 0 & 0 & 0 \\ \mathbf{C} & -\mathbf{I}_y & 0 & 0 & 0 & 0 \\ 0 & \mathbf{D} & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta s \\ \Delta\lambda \\ \Delta\mu \\ \Delta\gamma \end{bmatrix} = -\hat{r} \tag{6}$$

where we use the definitions ;

$$\begin{aligned}
H(x,\lambda) &= \nabla^2 f(x) + \mathcal{H}_g(x,\lambda) \ , \\
\nabla^2 f(x) &= \text{blkdiag}(\nabla^2 f^1(x), \dots, \nabla^2 f^m(x)) \ , \\
\mathcal{H}_g(x,\lambda) &= \text{blkdiag}(\sum_{j=1}^{n^1} \lambda_j^1 \nabla^2 g_j^1(x), \dots, \sum_{j=1}^{n^m} \lambda_j^m \nabla^2 g_j^m(x)) \ , \\
\mathcal{J}_g(x) &= \text{blkdiag}(\mathcal{J}_g^1(x^1), \dots, \mathcal{J}_g^m(x^m)) \ , \\
\mathcal{J}_g^i(x^i) &= \left[\nabla g_1^i(x^i), \dots, \nabla g_{n^i}^i(x^i)\right]^T \ , \\
\Lambda &= \text{blkdiag}(\Lambda^1, \dots, \Lambda^m), \ \Lambda^i = \text{diag}(\lambda^1, \dots, \lambda^m) \ , \\
S &= \text{blkdiag}(S^1, \dots, S^m), \ S^i = \text{diag}(s^1, \dots, s^m) \ .
\end{aligned}$$

'Long step' methods use the maximum stepsize that preserves strict positivity of the pair λ, s such that $(s + t\Delta s, \lambda + t\Delta\lambda) > 0$ holds, whereas 'short step' variants further restrict the step size to preserve 'centrality' and/or to ensure sufficient decrease in the residuals. The primal dual interior point method will terminate when the duality gap measure

$$\eta = \lambda^T s$$

and the primal dual residuals are reduced below desired thresholds. η becomes equal to the actual duality gap, when the iterates are primal and dual feasible, $r_{\text{dual}} = 0$, $r_{\text{primal}} = 0$ [6]. For a uniform perturbation vector β, when the perturbed KKT system is solved, that is we obtain a point on the central path, we have that

$$\hat{\eta} := \frac{\eta}{M} = \frac{\lambda^T s}{M}, \ \ \hat{\eta}\mathbf{1} = \beta \ ,$$

where $\lambda, s \in \mathbb{R}^M$, M is the total number of inequality constraints $M = \sum_i n^i$ and average complementarity value $\hat{\eta}$ provides a measure for the duality gap. In practice it is not necessary to wait for convergence to the central path, before reducing the perturbation parameter. In the standard 'feasible path following' method, for which convergence and complexity results are relatively easy to provide [19], the perturbation vector at each step is chosen as

$$\beta = \sigma\hat{\eta}\mathbf{1}, \quad \sigma \in (0,1) \ .$$

To ensure that proximity to the central path is preserved, the iterates can be constrained to stay within certain neighborhoods [8], [19].

B. Predictor-corrector method

Practical implementations of PDIP methods involve many heuristic modifications for accelerating the algorithm. Mehrotra's predictor corrector [15] together with Gondzia's multiple corrections [10] is widely used in commercial solvers [2] and results in much faster convergence. In this section we will briefly explain the predictor-corrector approach. For brevity, we write the Netwon system in compact form

$$\mathcal{M}\Delta z = -\begin{bmatrix} r_{\text{primal,dual}} \\ \Lambda S\mathbf{1} - \sigma\hat{\eta}\mathbf{1} \end{bmatrix}.$$

The Newton step for the original KKT system is given by

$$\mathcal{M}\Delta z^{\text{aff}} = -\begin{bmatrix} r_{\text{primal,dual}} \\ \Lambda S\mathbf{1} \end{bmatrix} \ ,$$

which is also known as the 'affine scaling direction' [19]. The duality gap reduction parameter σ is chosen according to the performance of the affine scaling direction. The duality gap measure that would be achieved by the affine scaling direction with the maximum feasible step-size t^{aff} is

$$\hat{\eta}^{\text{aff}} = (\lambda + t^{\text{aff}}\Delta\lambda^{\text{aff}})^T(s + t^{\text{aff}}\Delta s^{\text{aff}})/M \ .$$

The reduction parameter is chosen as $\sigma = \left(\hat{\eta}^{\text{aff}}/\hat{\eta}\right)^3$. Further information is extracted from the affine scaling step by computing a corrector term, which aims to correct the error in the linearization by adding a second order approximation. The composite predictor-corrector step is computed by solving

$$\mathcal{M}\Delta z^{\text{pc}} = -\begin{bmatrix} r_{\text{primal,dual}} \\ \Lambda S\mathbf{1} + v - \sigma\hat{\eta}\mathbf{1} \end{bmatrix} \ ,$$

where the corrector v is defined as

$$v = \rho\,(\Delta\Lambda)^{\text{aff}}\,(\Delta S)^{\text{aff}}\,\mathbf{1},$$

and ρ is a weighing factor for the corrector (for example, $\rho = t^{\text{aff}}$ [22] or $\rho = (t^{\text{aff}})^2$ [7]).

III. Coordinated Primal-Dual Interior point method

A. Coordinated Newton step

In this section we will show how to exploit the structure of the linear system (6), such that a Newton direction can be computed by communicating minimum amount of information between the agents and the central coordinator.

We start by eliminating $\Delta s, \Delta\lambda, \Delta x, \Delta\mu$ from the system, in the given order, to obtain

$$\begin{aligned}
\Delta s &= -\Lambda^{-1}S\Delta\lambda - \Lambda^{-1}r_{\text{cent}} \ , \\
\Delta\lambda &= S^{-1}\Lambda\tilde{r}_{\text{pr,in,x}} + S^{-1}\Lambda\mathcal{J}_g(x)\Delta x \ , \\
\Delta x &= -\mathbf{R}^{-1}\tilde{r}_{\text{dual},x} + -\mathbf{R}^{-1}\mathbf{C}^T\Delta\mu \ , \\
\Delta\mu &= -\mathbf{P}\Delta y + \mathbf{P}\tilde{r}_x \ ,
\end{aligned}$$

where we use the definitions

$$\begin{aligned}
\tilde{r}_{\text{prim,in},x} &= r_{\text{prim,in},x} - \Lambda^{-1}r_{\text{cent}} \ , \\
\tilde{r}_{\text{dual,x}} &= r_{\text{dual,x}} + \mathcal{J}_g(x)^T S^{-1}\Lambda\tilde{r}_{\text{prim,in,x}} \ , \\
\mathbf{R} &= H(x,\lambda) + \mathcal{J}_g(x)^T S^{-1}\Lambda\mathcal{J}_g(x) \ , \\
\tilde{r}_x &= r_{\text{prim,eq,x}} - \mathbf{C}\mathbf{R}^{-1}\tilde{r}_{\text{dual},x} \ , \\
\mathbf{P} &= (\mathbf{C}\mathbf{R}^{-1}\mathbf{C}^T)^{-1} \ , \\
q &= (r_{\text{dual},y} - \mathbf{P}\tilde{r}_x) \ .
\end{aligned}$$

Finally we are left with two equations

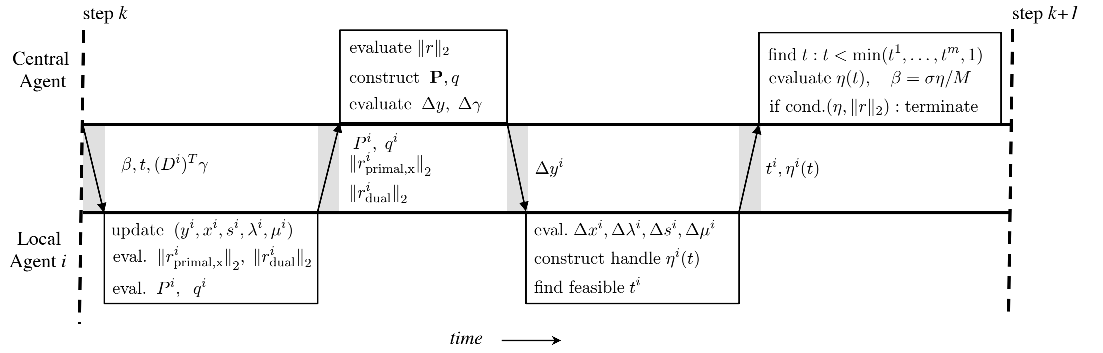
$$\begin{aligned}
\mathbf{P}\Delta y + \mathbf{D}^T\Delta\gamma &= -q \ , \\
\mathbf{D}\Delta y &= -r_{\text{prim},y} \ ,
\end{aligned}$$

which can be written in a compact form

$$\begin{bmatrix} \mathbf{P} & \mathbf{D}^T \\ \mathbf{D} & 0 \ , \end{bmatrix} \begin{bmatrix} \Delta y \\ \Delta\gamma \end{bmatrix} = -\begin{bmatrix} q \\ r_{\text{prim},y} \end{bmatrix} \ , \tag{7}$$

where

$$\mathbf{P} = \text{blkdiag}(P^1, \dots, P^m), q = ((q^1)^T, \dots, (q^m)^T)^T.$$



step k
step k+1
Central Agent
Local Agent i
evaluate $\|r\|_2$
construct $\mathbf{P}, q$
evaluate $\Delta y,\ \Delta\gamma$
find $t : t < \min(t^1, \ldots, t^m, 1)$
evaluate $\eta(t), \quad \beta = \sigma\eta/M$
if cond.$(\eta, \|r\|_2)$: terminate
$\beta, t, (D^i)^T\gamma$
$P^i,\ q^i$
$\|r^i_{\text{primal,x}}\|_2$
$\|r^i_{\text{dual}}\|_2$
Δy^i
$t^i, \eta^i(t)$
update $(y^i, x^i, s^i, \lambda^i, \mu^i)$
eval. $\|r^i_{\text{primal,x}}\|_2,\ \|r^i_{\text{dual}}\|_2$
eval. $P^i,\ q^i$
eval. $\Delta x^i, \Delta\lambda^i, \Delta s^i, \Delta\mu^i$
construct handle $\eta^i(t)$
find feasible t^i
time

Fig. 1: Newton step for the coordinated PDIP method

The matrices $\mathbf{R}$ and $\mathbf{P}$ are guaranteed to be invertible in virtue of the assumptions made in Section II. This system of equations can be solved in a coordinated manner as illustrated in Figure 1. The coordinated step distributes the linear algebra operations required to compute P^i and q^i to the agents, which can operate in parallel, and therefore reduces the computational effort at the central agent. Furthermore, local data of the agents is partially hidden, since the central agent does not have direct access to the information regarding the values of local variables $x^i, s^i, \mu^i, \lambda^i$ or functions $g^i(x^i), f^i(x^i)$.

Starting the iteration step k of the coordinated PDIP method, the agents receive the global step-size t, the perturbation (or barrier) parameter β and either the global dual variable γ or directly the product $(D^i)^T\gamma$, which will be used in the computation of $r_{\text{dual},y}$ and q^i. Firstly the agent updates its local variables and residuals using the step-size and the direction computed in the previous step. Using these values the agent then constructs the data P^i, q^i. The data (P^i, q^i) is sent to the central agent together with the norms of local residuals. The central agent constructs the overall residual norm, block diagonal $\mathbf{P}$ and stacked q and solves (7) for obtaining the direction in output variables Δy. Upon receiving it's local direction Δy^i the agent evaluates $\Delta x^i, \Delta\lambda^i, \Delta s^i$, the maximum feasible step-size and constructs the one-dimensional quadratic function $\eta^i(t)$ which outputs the sum of the local complementarity products for a given step-size t. Finally the central agent collects local maximum step-sizes t^i and the local 'duality gap' functions $\eta^i(t)$ for computing the global step-size t, evaluating the duality gap measure $\hat{\eta}$ and the perturbation vector β according to the reduction parameter σ. Having access to $\hat{\eta}$ and the norm of the primal-dual residuals of the previous step, the central agent can check if the specified termination criterion is satisfied.

Inspection of Figure 1 reveals that two communication rounds between the central agent and local agents are necessary to complete a single Newton step; one for computing parameters $\mathbf{P}, q$ and another for coordinating the step-sizes. In case of a short step approach, where the step-size is selected as to guarantee centrality and/or sufficient decrease of the residuals, additional communication steps might be necessary.

B. Decentralized predictor-corrector

In this section we will investigate how the predictor-corrector methodology can be applied to the coordinated PDIP setting, without introducing extra communication burden. The predictor-corrector approach requires computing two Newton steps per iteration, using the same Hessian. In practice, the reduction in the number of iterations greatly compensates the extra time spent in additional back-solves. However in the coordinated setting, the cost of communication becomes considerable, as now four communication rounds are necessary per iteration. In order to remedy this effect, we propose a decentralized predictor-corrector approach, which does not require additional communication rounds.

We break the communication in the predictor step and accept a cruder prediction; all agents solve a local affine scaling problem, disregarding the global equality constraint. This decentralized affine scaling step is used to determine the local duality gap reduction parameters σ^i and corrector vectors. The actual Newton step is then determined in a coordinated manner, respecting the coupling constraints. The main modification to the standard coordinated Newton step is in the first computational block shown in Figure 1, and changes the way local agents compute q^i using the local centrality residual r^i_{cent} as

$$
\begin{aligned}
\Delta(y^i)^{\text{aff}} &= -(P^i)^{-1} q^{i,\text{aff}} \\
(t^i)^{\text{aff}} &\ :\ (s^i + (t^i)^{\text{aff}}\Delta s^i, \lambda^i + (t^i)^{\text{aff}}\Delta\lambda^i) > 0 \\
\sigma^i &= \left(\hat{\eta}^{i,\text{aff}}/\hat{\eta}^i\right)^3 \\
v^i &= (\Delta\Lambda^i)^{\text{aff}}(\Delta S^i)^{\text{aff}}\mathbf{1} \\
r^i_{\text{cent}} &= \Lambda^i S^i \mathbf{1} + v^i - \sigma^i\hat{\eta}^i\mathbf{1},
\end{aligned}
$$

where q^{aff} is computed using the local complementarity residual r^i_{comp} as explained in Section II-B. The rest of the process is identical to the standard coordinated Newton step as shown in Figure 1, except that the global reduction parameter σ and the perturbation vector β are not needed.

Note that the local affine scaling direction could be significantly different from the original affine scaling direction, since the coupling constraint $\mathbf{D}\Delta y = -r_{\text{primal},y}$ is ignored. This diminishes the quality of the corrector and the reduction parameter σ^i in exchange for reduced communication effort.

In addition, by allowing different σ^i and corrector terms for each agent, we break the uniformity in the perturbation vector β. This last effect is not necessarily undesired, as discussed by [8] and may even enhance the robustness of the method. For guaranteed convergence in this setting, certain conditions that preserves centrality should be enforced, as explained in the next section.

C. Convergence with non-uniform perturbation

Consider the modified Newton step with a non-uniform perturbation vector on the complementarity condition:

$$\mathcal{M}\Delta z = -\begin{bmatrix} r_{\text{prim,dual}} \\ \Lambda S\mathbf{1} - \Lambda S\xi \end{bmatrix} \tag{8}$$

where the vector ξ specifies how much change in each complementarity pair is desired, with $\xi \in \mathbb{R}^M$ and $\xi > 0$. For a given $\sigma > 0$, ξ can be selected as $\xi_i = \sigma\frac{\hat{\eta}}{\lambda_i s_i}$ to recover the standard uniform perturbation vector with σ. This more generic definition captures the effect of the decentralized predictor-corrector described in the previous chapter. Note that any corrector applied together with the uniform-perturbation can be considered as a non-uniform perturbation vector.

In this section, we will state the convergence conditions for quadratic programs (QP) in the feasible case, where the primal-dual residuals are equal to 0, to make the analysis simpler as is common in the interior point literature [12], [23]. The results can be extended to the non-feasible case, and for LPs and QPs, a feasible point can be obtained with reformulations of the problem [11], [12].

We restrict the iterations to lie inside the 'symmetric neighborhood' [12],

$$\mathcal{N}_s = \{(x,y,s,\lambda) \in \mathcal{F}^0 : \frac{1}{\psi}\hat{\eta} \geq \lambda_j s_j \geq \psi\hat{\eta}, \quad \forall j \in [1,M]\}$$

for some $1 \gg \psi > 0$, where the strictly feasible set $\mathcal{F}^0$ is defined as $\mathcal{F}^0 = \{(x,y,s,\lambda) : (r_{\text{prim}}, r_{\text{dual}}) = 0,\ (s,\lambda) > 0\}$. Furthermore we restrict the reduction vector ξ to satisfy the following criteria,

$$\exists c\ : 0 < c < 1,\ \frac{\sum_j \xi_j\lambda_j s_j}{M} = c\hat{\eta} \tag{9}$$

$$\frac{1}{\phi}\frac{c\hat{\eta}}{\lambda_j s_j} \geq \xi_j \geq \phi\frac{c\hat{\eta}}{\lambda_j s_j} \tag{10}$$

for some $1 \gg \phi > \psi$. The constraints (10), (9) enforce that the next point we would like to arrive with the Newton step, does not have to be on the central path as in the uniform case, but it should be at least strictly within the $\mathcal{N}_s$ and aim to reduce the overall duality gap.

By applying these conditions, we will ensure that the linearization error in the Newton step is bounded, and there is always some finite step-size that can reduce the duality-gap and keep the system within $\mathcal{N}_s$. The following theorem states the convergence and complexity result.

Theorem 1: If problem (1) is a convex QP, given $\epsilon > 0$, suppose that a starting point $(x_0, y_0, s_0, \lambda_0) \in \mathcal{N}_s$ satisfies

$$\eta_0 \leq \frac{1}{\epsilon^\kappa} \tag{11}$$

for some positive constant κ. Let $\{\eta_k\}$ be a sequence generated by the iteration scheme, which takes steps using the Newton relation (8) with admissible parameters ϕ, ψ and reduction vectors ξ_k that satisfy (9), (10), with some $\alpha, \gamma \in (0,1), \gamma < 1-\alpha$ and $c_k \leq \gamma$. Then, there exists an admissible step-size $t_k \sim \mathcal{O}(\frac{1}{M})$ and an index $K \sim \mathcal{O}(M\log(\frac{1}{\epsilon}))$, such that

$$\eta_k \leq \epsilon,\ \ \forall k \geq K.$$

Proof: Due to space limitations, we refer the reader to the proof of Theorem 1 of [4] that covers generic convex QPs.

Integrating convergence conditions to the coordinated setting is not straightforward as one should balance extra communication burden and performance deterioration with the robustness of the algorithm. Here we state one possible way that does not require extra communication, but a detailed comparison and performance analysis is out of the scope of this paper. Constraints (10), (9) on ξ can be satisfied by enforcing $\frac{1}{\phi}c_{\text{min}}\hat{\eta} \geq \sigma^i\hat{\eta}^i\mathbf{1} - w^i v^i \geq \phi c_{\text{max}}\hat{\eta}$ and $c_{\text{max}} \geq (\sum_j \xi^i_j\lambda^i_j s^i_j)/M^i \geq c_{\text{min}}$ in the decentralized predictor-corrector computation step using pre-determined values $0 < c_{\text{min}} < c_{\text{max}} < 1$ and a corrector weighing term w^i that is selected by the agent. The central agent collects $\max(|\xi_i - 1|)$ from agents and constructs $\theta = 2^{-3/2}\|\xi - \mathbf{1}\|^2_\infty$ and bounding functions on the duality gap measure as $\hat{\eta}_{\text{max}}(t) := \hat{\eta}((1-t)+tc+t^2\theta)$ and $\hat{\eta}_{\text{min}}(t) := \hat{\eta}((1-t)+tc)$ which can be sent to the local agents together with Δy^i. The agents then compute their maximum allowable step-size that satisfy the neighborhood constraints as $t^i_{\text{max}}\ :\ (1/\psi)\hat{\eta}_{\text{min}}(t) \geq (s^i + t\Delta s^i)^T(\lambda^i + t^i\Delta\lambda^i) \geq \psi\hat{\eta}_{\text{max}}(t)$. Finally, the sufficient decrease condition can be verified by collecting local duality gap functions $\eta^i(t)$.

D. Experiments

We apply three different coordinated PDIP methods, PF: path following with a fixed reduction parameter σ and $\beta = \sigma\hat{\eta}\mathbf{1}$, PC: predictor-corrector, DPC: decentralized predictor corrector, on randomly generated problems as well as on a realistic demand response problem taken from [5]. In order to compare maximum performance, we omit convergence safeguards and use the 'long-step' approach for step-size selection, which is a common practice [19]. On randomly generated problems, the benefit of the decentralized predictors in terms of reduced communication burden is clear as shown in Figure 2. However this reduction is not deterministic, as can be seen in Figure 3, where the performance of PF method and DPC method are similar. In terms of iterations, PC method prevails, but suffers from increased communication rounds. The DPC method is effective in reducing the communication burden, however the reduction depends on the problem data and the starting point. The standard path-following method is also attractive due to small communication per iteration requirement and simplicity of the algorithm.

E. Conclusion

In this paper, we have discussed three different coordinated PDIP methods, standard path following, Mehrotra's predictor-corrector and the novel decentralized predictor-corrector method, for solving the multi-agent problem (1). We provide convergence results for PDIP methods with non-uniform complementarity perturbations, as the decentralized PC scheme falls into this category. This convergence result can be applied to any PDIP method that utilizes corrector terms and breaks the uniformity of the complementarity perturbation. Finally we have tested the coordinated PDIP methods on numerical examples. According to the results, depending on the expense of iterations and communications, all three methods are practical. Future work will include integrating the convergence safeguards in an optimal manner to the decentralized predictor-corrector scheme as well as generic predictor corrector schemes and more in-depth comparison on realistic multi-agent optimization problems.

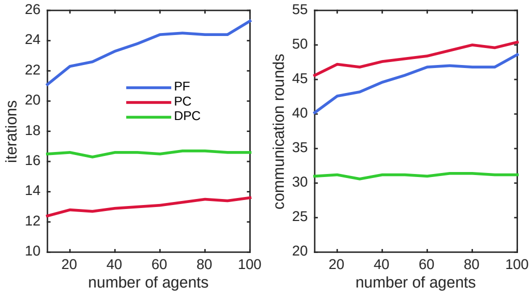


Fig. 2: Evolution of iteration and communication rounds required to bring the suboptimality $\|y - y^*\|/n_y$ below 10^{-5}, for increasing number of agents for randomly generated coupled QPs with feasible starting points, using the mean of 10 sampled problems with $n_x = 20$, $n_y = 5$.

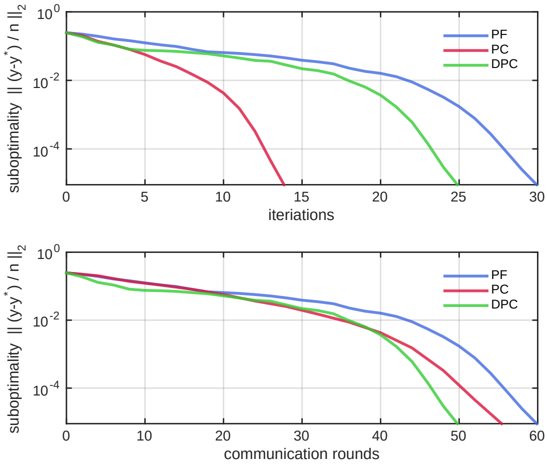


Fig. 3: Comparison of various CPDIP mehtods for a realistic example of demand response [5]. In this case 12 smart-buildings help a distribution grid for voltage support in case of a significant solar power in feed. The algorithm is started from a point that is locally centered but infeasible with respect to the coupling constraint.

References

[1] M. Annergren, S. K. Pakazad, A. Hansson, and B. Wahlberg. A distributed primal-dual interior-point method for loosely coupled problems using ADMM. *arXiv:1406.2192 [math]*, June 2014. arXiv: 1406.2192.

[2] H. Y. Benson, J. J. Cochran, L. A. Cox, P. Keskinocak, J. P. Kharoufeh, and J. C. Smith. Interior-Point Linear Programming Solvers. In *Wiley Encyclopedia of Operations Research and Management Science*. John Wiley & Sons, Inc., 2010.

[3] D. Bertsekas. *Nonlinear Optimization*. Athena Scientific, 2008.

[4] A. Bitlislioglu and C. N. Jones. On generalized primal-dual interior-point methods with non-uniform complementarity perturbations for quadratic programming. Technical Report 231144, École Polytechnique Fédérale de Lausanne, September 2017.

[5] A. Bitlislioglu, I. Pejcic, and C. N. Jones. Interior point decomposition for multi-agent optimization. In *IFAC 2017 World Congress, Toulouse, France*, July 2017.

[6] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, Cambridge, 2004.

[7] C. Cartis. On the convergence of a primal-dual second-order corrector interior point algorithm for linear programming. Technical Report 05/04, Numerical Analysis Group, Computing Laboratory, Oxford University, March 2005.

[8] M. Colombo and J. Gondzio. Further development of multiple centrality correctors for interior point methods. *Computational Optimization and Applications*, 41(3):277–305, December 2008.

[9] A. Domahidi, E. Chu, and S. Boyd. ECOS: An SOCP solver for embedded systems. In *2013 European Control Conference (ECC)*, pages 3071–3076, July 2013.

[10] J. Gondzio. Multiple centrality corrections in a primal-dual method for linear programming. *Computational Optimization and Applications*, 6(2):137–156, September 1996.

[11] J. Gondzio. Interior point methods 25 years later. *European Journal of Operational Research*, 218(3):587–601, May 2012.

[12] J. Gondzio. Convergence Analysis of an Inexact Feasible Interior Point Method for Convex Quadratic Programming. *SIAM Journal on Optimization*, 23(3):1510–1527, January 2013.

[13] E. Klintberg and S. Gros. A primal-dual Newton method for distributed Quadratic Programming. In *53rd IEEE Conference on Decision and Control*, pages 5843–5848, December 2014.

[14] J. Mattingley and S. Boyd. CVXGEN: a code generator for embedded convex optimization. *Optimization and Engineering*, 13(1):1–27, March 2012.

[15] S. Mehrotra. On the Implementation of a Primal-Dual Interior Point Method. *SIAM Journal on Optimization*, 2(4):575–601, November 1992.

[16] S. Mehrotra and M. G. Ozevin. Decomposition Based Interior Point Methods for Two-Stage Stochastic Convex Quadratic Programs with Recourse. *Operations Research*, 57(4):964–974, May 2009.

[17] I. Necoara, V. Nedelcu, and I. Dumitrache. Parallel and distributed optimization methods for estimation and control in networks. *Journal of Process Control*, 21(5):756–766, June 2011.

[18] I. Necoara and J. A. K. Suykens. Interior-Point Lagrangian Decomposition Method for Separable Convex Optimization. *Journal of Optimization Theory and Applications*, 143(3):567, May 2009.

[19] J. Nocedal and J. Wright, S. *Numerical Optimization*. Springer, New York, 2000.

[20] S. K. Pakazad, A. Hansson, M. S. Andersen, and I. Nielsen. Distributed primaldual interior-point methods for solving tree-structured coupled convex problems using message-passing. *Optimization Methods and Software*, 32(3):401–435, May 2017.

[21] F. A. Potra and S. J. Wright. Interior-point methods. *Journal of Computational and Applied Mathematics*, 124(12):281–302, December 2000.

[22] M. Salahi, J. Peng, and T. Terlaky. On Mehrotra-Type Predictor-Corrector Algorithms. *SIAM Journal on Optimization*, 18(4):1377–1397, December 2007.

[23] S. J. Wright. *Primal-Dual Interior-Point Methods*. SIAM, Philadelphia, 1997.

[24] G. Zhao. A Log-Barrier method with Benders decomposition for solving two-stage stochastic linear programs. *Mathematical Programming*, 90(3):507–536, May 2001.

[25] G. Zhao. A Lagrangian Dual Method with Self-Concordant Barriers for Multi-Stage Stochastic Convex Programming. *Mathematical Programming*, 102(1):1–24, January 2005.