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. We analyse the general framework of PDIP methods with non-uniform perturbations and provide convergence and complexity results, that directly apply to the proposed coordinated PDIP with decentralized predictor-corrector scheme.

I. INTRODUCTION

We consider the following convex multi-agent optimization problem:

\[ \begin{align*}
\min_{x^1, \ldots, x^m} & \sum_{i=1}^{M} f^i(x^i) \\
\text{s.t.} & \quad g_i(x^i) \leq 0 \\
& \quad C^i x^i = y^i \\
& \quad \sum_{i=1}^{m} D^i y^i = d
\end{align*} \]

(1)

where subscript \( i \) indicates that the function or variable belongs to agent \( i \). The cost function and the constraint set are separable, except 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.

The 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). Furthermore it can capture coupling in the cost and inequality constraints 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 or 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 in the information network. Many engineering applications, such as the power grid 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 the 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 first order methods such as sub-gradient or cutting plane methods [2]. 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 [3], [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 [3], [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 [20], where an effort was also made to distribute the Newton step to eliminate the need for a central agent, albeit in a restricted setting of ‘loosely coupled’ problems.
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 [9], [15], PDIP methods can become very efficient and therefore are commonly used in implementations aiming for high-performance [1], [8], [14]. However, in the interior-point decomposition framework, none of the aforementioned works apply a predictor-corrector scheme for acceleration, 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 decomposition setting, with primal-dual iterations in the local agents. For brevity we re-write (1) in a compact form with slack variables

\[
\begin{align*}
\min_{x} & \quad f(x) \\
\text{s. t.} & \quad g(x) + s = 0 \\
& \quad Cx - y = 0 \quad (\mu) \\
& \quad Dy = d \quad (\gamma) \\
& \quad s \geq 0 \quad (\lambda)
\end{align*}
\]

where the dual variables \(\mu, \gamma, \lambda\) are shown next to their corresponding constraints. The first order optimality conditions for problem (2), or the KKT system, can be written as

\[
\begin{align*}
r_{\text{dual, } x}(x, \mu, \lambda) &= \nabla f(x) + C^T \mu + J_g(x)^T \lambda \\
r_{\text{dual, } y}(y, \mu) &= D^T \gamma - \mu \\
r_{\text{prim, in, } x}(x, s) &= g(x) + s \\
r_{\text{prim, eq, } x}(x, y) &= Cy - y \\
r_{\text{prim, y}}(y) &= Dy - d \\
r_{\text{comp}}(\lambda, s) &= \Lambda s \mathbf{1}
\end{align*}
\]

or in a compact form:

\[
r(x, y, s, \mu, \gamma, \lambda) = 0.
\]

Where \(r = (r_{\text{dual, } x}, r_{\text{comp}})\). PDIP methods operate by applying Newton’s method on a smoothed version of the nonlinear set of equations (4). The smoothing is obtained by perturbing the complementary slackness condition.

\[
r_{\text{comp}}(\beta, \lambda, s) = \Lambda s \mathbf{1} - \beta, \quad \beta > 0
\]

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 \(\beta\) is a scalar multiplied by a vector of ones. However we will let it attain different values for each constraint, similar to [7], 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:

\[
\begin{bmatrix}
H(x, \lambda) & 0 & 0 & J_g(x)^T & C^T & 0 \\
0 & 0 & 0 & 0 & -I_\lambda & D^T \\
0 & 0 & \Lambda & S & 0 & 0 \\
J_g(x) & 0 & I_s & 0 & 0 & 0 \\
C & -I_y & 0 & 0 & 0 & 0 \\
0 & D & 0 & 0 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
\Delta x \\
\Delta y \\
\Delta s \\
\Delta \lambda \\
\Delta \mu \\
\Delta \gamma
\end{bmatrix} = -r
\]

‘Long step’ methods use the maximum stepsizes that preserves strict positivity of the pair \(\lambda, s\) 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. \(\eta\) becomes equal to the actual duality gap, when the iterates are primal and dual feasible, \(r_{\text{dual}} = 0, r_{\text{prim}} = 0\) [5]. For a fixed scalar \(\beta\), when the perturbed KKT system is solved, that is we obtain a point on the central path, we have that

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

$$\beta = \sigma \bar{\eta}, \quad \sigma \in (0, 1)$$

To ensure that proximity to the central path is preserved, the iterates can be constrained to stay within certain neighborhoods [7], [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 Gondzía’s multiple corrections [9] is widely used in commercial solvers [1] and results in much faster convergence. In this section we will briefly explain the predictor-corrector approach. For brevity, we write the Newton system in compact form

$$H \Delta z = - \begin{bmatrix} r_{\text{prim,dual}} \\ \Lambda S \bar{1} - \sigma \bar{\eta} \bar{1} \end{bmatrix}$$

and we define the ‘affine scaling direction’ as

$$H \Delta z^{\text{aff}} = - \begin{bmatrix} r_{\text{prim,dual}} \\ \Lambda S \bar{1} \end{bmatrix}$$

which gives a Newton step for the original KKT system. The duality gap reduction parameter $\sigma$ 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 $t^{\text{eff}}$ is

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

The reduction parameter is chosen as $\sigma = (\bar{\eta}^{\text{aff}} / \bar{\eta})^3$. 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

$$H \Delta z^{\text{PC}} = - \begin{bmatrix} r_{\text{prim,dual}} \\ \Lambda S \bar{1} + v - \sigma \bar{\eta} \bar{1} \end{bmatrix}$$

where the corrector $v$ is defined as

$$v = \rho \Delta \lambda^{\text{aff}} \Delta s^{\text{aff}} \bar{1}$$

and $\rho$ is a weighing factor for the corrector (for example, $\rho = t^{\text{eff}}$ [22] or $\rho = (t^{\text{eff}})^2$ [6]).

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 $\Delta x, \Delta s, \Delta \lambda, \Delta \mu$ from the system (see Appendix). We are left with a more compact form

$$\begin{bmatrix} P & D^T \\ 0 & \bar{q} \end{bmatrix} \Delta \gamma = - \begin{bmatrix} q \\ \gamma_{\text{prim,y}} \end{bmatrix}$$

where

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

The linear system (7) can also be written as

$$P^i \Delta y^i + (D^i)^T \Delta \gamma = -q^i, \quad \forall i \in [1, \ldots, m]$$

revealing the underlying structure. 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 hidden, since the central agent does not have access to any information regarding the values of local variables $x^i, s^i, \lambda^i, \mu^i$ or functions $q^i(x^i), f^i(x^i)$.

Starting the iteration step $k$ of the coordinated PDIP method, the agents receive the global step-size $t$, the perturbation (or barrier) parameter $\beta$ and either the global dual variable $\gamma$ or directly the product $(D^k)^T \gamma$, which will be used in the computation of $r^1_{\text{dual,y}}$ and $q^i$. Firstly the agent updates it’s 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 $P^i, q^i$ (see Appendix). 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 $P$ and stacked $q$ and solves (7) for obtaining the direction in output variables $\Delta \gamma$. Upon receiving it’s local direction $\Delta y^i$ the agent evaluates $\Delta x^i, \Delta \lambda^i, \Delta s^i$ (see Appendix), the maximum feasible step-size and constructs the 1 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 $\bar{\eta}$ and the perturbation vector $\beta$ according to the reduction parameter $\sigma$. Having access to $\bar{\eta}$ 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 $P, q$ 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 two linear system solves with the same left-hand side matrix. 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 $\sigma^t$ 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 $q^t$ using the local centrality residual $r^\text{cent}$ as

$$\Delta y^t = -(P^t)^{-1}q^{t,\text{aff}}$$

$$(t^i)^\text{aff}: (s^t + (t^i)^\text{aff}\Delta s^t, \lambda^t + (t^i)^\text{aff}\Delta \lambda^t) > 0$$

$$\sigma^t = (\hat{\eta}^{i,\text{aff}}/\hat{\eta}^t)^{3/2}$$

$$v^t = (\Delta \lambda^t)^{\text{aff}}(\Delta S^t)^{\text{aff}}1 + v^t - \sigma^t \hat{\eta}^t 1$$

where $q^{\text{aff}}$ is computed using the local complementarity residual $r^\text{comp}$ 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 $\sigma$ and the perturbation vector $\beta$ are not needed.

Note that the local affine scaling direction could be significantly different from the original affine scaling direction, since the coupling constraint $D\Delta y = -r_{\text{prim,} y}$ is ignored. This diminishes the quality of the corrector and the reduction parameter $\sigma^t$ in exchange of reduced communication effort. In addition, by allowing different $\sigma^t$ and corrector terms for each agent, we break the uniformity in the perturbation vector $\beta$. This last effect is not necessarily a bad thing, as discussed by [7] 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:

$$H \Delta z = - \left[ \begin{array}{c} r_{\text{prim, dual}} \\ \Delta S 1 - \Delta S \xi \end{array} \right]$$

where the vector $\xi$ specifies how much reduction in each complementarity pair is desired, with $\xi \in \mathbb{R}^M$ and $\xi > 0$. For a given $\sigma > 0$, $\xi$ can be selected as $\xi_j = \sigma_j \lambda^t_j$ to recover the standard uniform perturbation vector with $\sigma$. 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 to make the analysis simpler, as is common in the interior point literature [10], [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 [10], [11].

We restrict the iterations to lie inside the 'symmetric neighborhood' [10],

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

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 $\xi$ to satisfy the following criteria,
\begin{equation}
\exists c : 0 < c < 1, \sum_j \lambda_j s_j = \frac{c \bar{\eta}}{M} \tag{10}
\end{equation}
\begin{equation}
\frac{1}{\phi} \frac{c \bar{\eta}}{\lambda_j s_j} \geq \xi_j \geq \frac{c \bar{\eta}}{\lambda_j s_j} \tag{11}
\end{equation}
for some $1 \gg \phi > \psi$. The constraints (11), (10) 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 the 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
\begin{equation}
\eta_0 \leq \frac{1}{\epsilon^\alpha} \tag{12}
\end{equation}
for some positive constant $\alpha$. Let $\{\eta_k\}$ be a sequence generated by the iteration scheme, which takes steps using the Newton relation (8) with admissible parameters $\phi, \psi$ and reduction vectors $\xi_k$ that satisfy (10), (11), with some $\alpha, \gamma \in (0,1), \gamma < 1 - \alpha$ and $c_k \geq \gamma$. Then, there exists a step-size $\tau_k \sim O(\frac{1}{\epsilon^\alpha})$ such that
\begin{equation}
\eta_k \leq \epsilon, \forall k \geq K.
\end{equation}

**Proof:** see Appendix.

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 (11), (10) on $\xi$ can be satisfied by enforcing
\begin{equation}
\begin{aligned}
\frac{1}{\phi} \frac{c_{\min} \bar{\eta}}{\lambda_j s_j} &\geq \alpha + \psi t^2 + 1 - w^2 t^2 \\
c_{\max} &\geq (\sum_j \frac{\lambda_j s_j}{M}) / \epsilon^\alpha \geq c_{\min}
\end{aligned}
\end{equation}
in the decentralized predictor-corrector computation step using pre-determined values $0 < c_{\min} < c_{\max} < 1$ and a corrector weighing term $w^2$ that is selected by the agent. The central agent collects max $|\xi_i - 1|$ from agents and constructs $\theta = 2^{-3/2}\|\xi - 1\|_2^2$ and bounding functions on the duality gap measure as
\begin{equation}
\bar{\eta}_{\max}(t) := \bar{\eta}(1 - t) + tc + t^2 \theta \\
\bar{\eta}_{\min}(t) := \bar{\eta}(1 - t) + tc
\end{equation}
which can be sent to the local agents together with $\Delta \eta^i$. The agents then compute their maximum allowable step-size that satisfy the neighborhood constraints as;

\begin{equation}
\tau_{\max} := (1/\psi) \bar{\eta}_{\min}(t) \geq (\psi + t^2 \Delta \lambda^i)^2 (\lambda^i + t^2 \Delta \lambda^i) \geq \psi \bar{\eta}_{\max}(t)
\end{equation}
Finally, the sufficient decrease condition can be verified at the central node using the collected local duality gap functions $\bar{\eta}^i(t)$.

**D. Experiments**

We apply three different coordinated PDIP methods, PF: path following with a fixed reduction parameter $\eta$ and $\beta = \sigma \bar{\eta}$, PC: predictor-corrector, DPC: decentralized predictor corrector, on randomly generated problems as well as on a realistic demand response problem taken from [3]. 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, Mehrota’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.

**REFERENCES**


with respect to the coupling constraint. started from a point that is locally centered but infeasible in case of a significant solar power infeed. The algorithm is build from buildings (modelled with OpenBuild [12]) help a distribution example of demand response [3]. In this case 12 smart-coupled QPs with feasible starting points, using the mean of 10 sampled problems with $n_x = 20$, $n_y = 5$.

![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 $n_x = 20$, $n_y = 5$.](image)

![Fig. 3: Comparison of various CPDIP methods for a realistic example of demand response [3]. In this case 12 smart-buildings (modelled with OpenBuild [12]) help a distribution grid (simplified IEEE 123 test feeder [4]) for voltage support in case of a significant solar power infed. The algorithm is started from a point that is locally centered but infeasible with respect to the coupling constraint.](image)

**APPENDIX**

### A. Definitions

Compact multi-agent problem (2):

$$s = [ (s^1)^T \ldots (s^n)^T ]^T$$

$$f(x) = \sum_{i=1}^m f_i(x^i), g(x) = [g^1(x^1)^T \ldots g^m(x^m)^T]^T$$

$$C = \text{blkdiag}(C^1, \ldots , C^m), D = [ D^1 \ldots D^m ]$$

$y, x, d, \mu, \lambda$ are defined similar to $s$.

Primal dual Newton system (6):

$$H(x, \lambda) = \nabla^2 f(x) + H_d(x, \lambda)$$

$$\nabla^2 f(x) = \text{blkdiag}(\nabla^2 f_1(x), \ldots , \nabla^2 f_m(x))$$

$$H_d(x, \lambda) = \text{blkdiag}(\sum_{j=1}^{n_1} \lambda^j \nabla^2 g^j_1(x), \ldots , \sum_{j=1}^{n_m} \lambda^j \nabla^2 g^m_j(x))$$

$$J_a(x) = \text{blkdiag}(J_a^1(x), \ldots , J_a^m(x))$$

$$J_b(x) = [ \nabla g^1_1(x), \ldots , \nabla g^m_{m_1}(x^m) ]^T$$

$$\Lambda = \text{blkdiag}(\Lambda^1, \ldots , \Lambda^m), \Lambda^j = \text{diag}(\lambda^j_1, \ldots , \lambda^j_m)$$

$$S = \text{blkdiag}(S^1, \ldots , S^m), S^i = \text{diag}(s^1, \ldots , s^m)$$

Note that $\Lambda, S > 0$, $H(x, \lambda) \succeq 0$. 

---


B. Variable elimination

Starting from the Newton system (6) one can the eliminate variables $\Delta s$, $\Delta \lambda$, $\Delta x$, $\Delta \mu$, in the given order, to obtain
\[
\begin{align*}
\Delta s &= -\Lambda^{-1}S\Delta \lambda - \Lambda^{-1}r_{\text{cent}} \\
\Delta \lambda &= S^{-1}A_{\text{prim.in},x} + S^{-1}A_J(y)\Delta x \\
\Delta x &= -R^{-1}\tilde{r}_{\text{dual},x} + -R^{-1}C^T\Delta \mu \\
\Delta \mu &= -P\Delta y + P\tilde{r}_x
\end{align*}
\]
where we use the definitions
\[
\begin{align*}
\tilde{r}_{\text{prim,in},x} &= r_{\text{prim,in},x} - \Lambda^{-1}r_{\text{cent}} \\
\tilde{r}_{\text{dual},x} &= r_{\text{dual},x} + J_g(x)T S^{-1}A_{\text{prim.in},x} \\
R &= H(x, \lambda) + J_g(x)^T S^{-1}A_J(y) \\
\tilde{r}_x &= r_{\text{prim.eq},x} - CR^{-1}\tilde{r}_{\text{dual},x} \\
P &= (CR^{-1}C^T)^{-1} \\
q &= (r_{\text{dual},y} - P\tilde{r}_x)
\end{align*}
\]
Finally we are left with two equations
\[
\begin{align*}
P\Delta y + D^T\Delta y &= -q \\
D\Delta y &= -r_{\text{prim},y}
\end{align*}
\]
which corresponds to (7).

C. Convergence proof

The analysis and the proof will be similar to that of [10] by showing that starting from a point within $N$, it is possible to keep the linearization error small and obtain an admissible step-size of complexity $O\left(\frac{1}{\mu^2}\right)$ that reduces the duality gap sufficiently and keep the iterates within $N$. Allowing non-uniformity necessitates additional constraints (11), (10) on the perturbation vector and results in different bounding terms compared to [10], however the complexity result is identical.

We start the analysis by deriving bounds on the linearization error.

**Lemma 1:** For $(x, y, s, \lambda) \in N$, the Newton direction resulting from (8) satisfies the following relations;
\[
\begin{align*}
0 &\leq \Delta x^T S x \leq \theta x^T S x \\
-\theta &\lambda^T S x \leq \Delta \lambda x \leq \sqrt{\theta} \lambda x
\end{align*}
\]
where $\theta$ is defined as
\[
\theta := 2^{-3/2} \max\left(\frac{c}{\phi \psi} - 1, 1 - \phi \psi\right).
\]

**Proof:** Using the Newton step equations (6) in the feasible setting, it can be shown that
\[
\Delta x^T S x = \Delta x^T H(x, \lambda) \Delta x
\]
Since $f_i$ and $g_i$’s are convex functions and $\lambda \geq 0$, we have $H(x, \lambda) \geq 0$, which validates the left-hand side inequality in (13).

For the right-hand side of (13) we need a technical result ([11], Lemma 3.3). Given two vectors $u, v \in \mathbb{R}^n$, if $u^T v \geq 0$ then,
\[
\|U V 1\| \leq 2^{-3/2} \|u + v\|^2
\]
where $U := \text{diag}(u)$, $V := \text{diag}(v)$. With relation (16) at hand, we start with the non-uniformly perturbed centrality equation
\[
S\Delta \lambda + \Delta S x = \Lambda S(\xi - 1).
\]
Multiply both sides with $(\Lambda S)^{-1/2}$;
\[
\Lambda^{-1/2} S^{1/2} S^1 \Delta \lambda + \Lambda^{1/2} S^{1/2} S^1 \Delta x = (\Lambda S)^{1/2}(\xi - 1).
\]
Since $\Delta \lambda, \lambda, S, \Delta S$ are all diagonal matrices we have that
\[
\|\Delta \lambda \Delta S \Delta 1\| = \|\Lambda^{-1/2} S^{1/2} \Delta \lambda \Lambda^{1/2} S^{-1/2} S^1 \Delta x\|
\]
Next we use (16) with $u := \Lambda^{-1/2} S^{1/2} \Delta \lambda$ and $v := \Lambda^{1/2} S^{-1/2} S^1 \Delta x$. Note that the condition $u^T v \geq 0$ is satisfied since $\Delta \lambda^T S \Delta x \geq 0$.

Finally using Cauchy-Schwarz inequality we arrive at the upper bound on $\Delta \lambda^T S \Delta x$;
\[
\Delta \lambda^T S \Delta x = (\Delta \lambda \Delta S \Delta 1)^T 1 \leq \|\Delta \lambda \Delta S \Delta 1\| \|1\| = \theta \lambda^T S
\]
which proves the right-hand side of (13).

Next, we prove the relation (14) that provide bounds on $\Delta \lambda S \Delta x$. We write equation (18) element-wise and square both sides
\[
\lambda_j^{1/2} s_j^{1/2} \Delta \lambda_j + \lambda_j^{1/2} s_j^{-1/2} \Delta s_j = \lambda_j^{1/2} s_j^{1/2}(\xi_j - 1) \\
\Rightarrow \lambda_j^{-1} s_j \Delta \lambda_j^2 + \lambda_j s_j^{-1} \Delta s_j^2 + 2\Delta \lambda_j \Delta s_j = (\xi_j - 1)^2 \lambda_j s_j
\]
Since $\lambda, s > 0$, we have that
\[
2\Delta \lambda_j \Delta s_j \leq (\xi_j - 1)^2 \lambda_j s_j
\]
Using the relation $\|\Delta \lambda \Delta S \Delta 1\| \leq \|\Delta \lambda \Delta S \Delta 1\| \leq \|\Delta \lambda \Delta S \Delta 1\|$ and the definition of $\theta$, we arrive at
\[
-\theta \lambda^T S \leq \Delta \lambda \Delta s_j \leq \sqrt{\theta} \lambda s_j
\]
In the following, we will show that a finite step-size can be obtained such that the neighborhood constraints are preserved and a sufficient decrease in the duality gap is achieved.
For a given direction, we describe the values after a step with a step-size \( t \in [0, 1] \) as

\[
\lambda_j(t) = \lambda_j + t \Delta \lambda_j, \quad s_j(t) = s_j + t \Delta s_j.
\]

**Lemma 2:** For a given point \( (x, y, s, \lambda) \in \mathcal{N}_s \), and the Newton direction resulting from (8), if the step-size \( t \in (0,1) \) satisfies the following conditions:

\[
t \leq \frac{c(\phi - \psi)}{(\psi + M)\theta} \quad (19)
\]

\[
t \leq \frac{c(\phi - \psi)}{\sqrt{2} \theta} \quad (20)
\]

then the new point attained by taking a step in the Newton direction with step-size \( t \) stays in the neighborhood \( \mathcal{N}_s \), that is

\[
(x(t), y(t), s(t), \lambda(t)) \in \mathcal{N}_s. \quad (21)
\]

**Proof:** Symmetric neighborhood \( \mathcal{N}_s \) implies lower and upper bound constraints on complementarity pairs and feasibility of the iterates. We start with the lower bound constraint. Using Lemma 1, the centrality equation (17) and neighborhood constraints on the current point \( \lambda_j s_j \geq \psi \hat{\eta} \) and the perturbation vector \( \xi_j \geq \phi c \) we obtain an initial lower bound on \( \lambda_j(t)s_j(t) \)

\[
\lambda_j(t)s_j(t) = (\lambda_j + t\Delta \lambda_j)(s_j + t\Delta s_j)
\]

\[
= \lambda_j s_j(1-t) + t\xi_j \lambda_j s_j + t^2 \Delta \lambda_j \Delta s_j
\]

\[
\geq \psi \hat{\eta}(1-t) + t\xi(c) - t^2 \theta M \hat{\eta}
\]

Next, we derive an upper bound on \( \tilde{\eta}(t) \), using the relation (10), centrality equation (17) and Lemma 1

\[
\tilde{\eta}(t) = \lambda(t)T s(t)/M
\]

\[
= (\lambda^T s + t(c - 1)\lambda^T s + t^2 \Delta \lambda^T \Delta s)/M
\]

\[
\leq (\lambda^T s + t(c - 1)\lambda^T s + t^2 \theta \lambda^T s)/M
\]

\[
= \hat{\eta} ((1-t) + tc + t^2 \theta)
\]

Using the upper bound on \( \hat{\eta}(t) \) and lower bound on \( \lambda_j(t)s_j(t) \) we can write a sufficient condition for \( \lambda_j(t)s_j(t) \geq \psi \hat{\eta}(t) \),

\[
\psi \hat{\eta}(1-t) + t \xi(c) - t^2 \theta M \hat{\eta} \geq \psi \hat{\eta}(1-t) + t \xi(c) + t^2 \theta \hat{\eta}
\]

\[
(\phi - \psi)\xi(c) \geq t ((\psi + M)\theta) \hat{\eta}
\]

which is identical to condition (19).

Next, we tackle the upper bound constraint on \( \lambda_j(t)s_j(t) \leq \frac{1}{\psi} \tilde{\eta}(t) \) which is equivalent to

\[
\lambda_j \lambda_j s_j(1-t) + t\xi_j \lambda_j s_j + t^2 \Delta \lambda_j \Delta s_j \leq \frac{1}{\psi} \tilde{\eta}(1-t) + t\xi(c) + t^2 \Delta \lambda^T \Delta s/M
\]

Seeing that \( \lambda_j s_j \leq \frac{1}{\psi} \tilde{\eta} \), and \( \xi_j \leq \frac{1}{\psi} \xi(c) \) a sufficient condition is

\[
\frac{1}{\psi} \tilde{\eta}(1-t) + t \frac{1}{\psi} \xi(c) + t^2 \Delta \lambda^T \Delta s/M
\]

\[
t \leq \frac{1}{\psi} c \hat{\eta} + t^2 \Delta \lambda_j \Delta s_j \leq \frac{1}{\psi} c \hat{\eta} + t^2 \Delta \lambda^T \Delta s/M
\]

\[
t^2 \left( \Delta \lambda_j \Delta s_j - \frac{1}{\psi} \Delta \lambda^T \Delta s/n \right) \leq \frac{te \hat{\eta}(1 - \frac{1}{\psi})}{\sqrt{2} \theta}
\]

Since \( \Delta \lambda_j \Delta s_j \leq \sqrt{2} \theta \hat{\eta} \),\( \Delta \lambda_j \Delta s_j \geq 0 \) another sufficient condition is

\[
t \sqrt{2} \theta \hat{\eta} \leq c \hat{\eta}(\frac{1}{\psi} - \frac{1}{\psi})
\]

\[
t \leq \frac{\psi c(\frac{1}{\psi} - \frac{1}{\psi})}{\sqrt{2} \theta} = \frac{c(\phi - \psi)}{\sqrt{2} \theta}
\]

which is equivalent to (20).

For a LP and QP, the KKT system is composed of linear equations, except for the complementarity equation. Therefore the Newton step with \( r_{\text{dual}} = 0, r_{\text{primal}} = 0 \), will preserve the feasibility for any step-size, from which we conclude that a step-size  \( t \in (0,1) \) satisfying (19), (20) will keep the iterates within \( \mathcal{N}_s \).

**Lemma 3:** For a given point \( (x, y, s, \lambda) \in \mathcal{N}_s \), parameter \( \alpha \in (0,1) \) and the Newton direction resulting from (8), if the step-size \( t \), and the mean reduction parameter \( c \), satisfy the conditions;

\[
1 - c > \alpha, \quad 0 < t \leq \frac{1 - c - \alpha}{\theta}
\]

then the new duality gap \( \eta(t) \) satisfies the reduction requirement

\[
\eta(t) \leq (1 - \alpha t) \eta .
\]

**Proof:**

From (22), we have that

\[
\eta(t) \leq (1 - \gamma(t)) \eta
\]

with \( \gamma(t) = t(1 - c) - t^2 \theta \). Therefore, a sufficient condition for satisfying (24) is

\[
\gamma(t) = t(1 - c) - t^2 \theta \geq \alpha t
\]

\[
\Rightarrow t^2 \theta \leq t(1 - c - \alpha) \Rightarrow 0 < t \leq \frac{1 - c - \alpha}{\theta}
\]

which is equivalent to (23).

In order to obtain the convergence and complexity result, first we will show that there exists a \( t \sim O(\frac{1}{M}) \) that satisfies (20), (19) and (23). Then, application of Theorem 3.2 from [23] will provide the desired result.

The set of admissible step-sizes are given by:

\[
\{ t \mid 0 < t \leq \min \left( \frac{1 - c - \alpha}{\theta}, \frac{c(\phi - \psi)}{(\psi + M)\theta} \right) \}
\]

(25)

where \( \theta := 2 - 3/\max (\frac{c}{\psi}, -1, \phi \psi c) \).

Choose \( \alpha = 0.1, \psi = 0.1 \) and \( \phi = 0.2 \) and fix reduction parameter \( c = 0.8 \) for every step. For these values \( \theta \) is given by \( 2^{-3/\max (0.2, 0.8 - 1)} \approx 13.8 \) It can be shown that the step-size selection \( \frac{1}{20M} \) will be always admissible, which provides \( t \sim O(\frac{1}{M}) \).

a) **Proof of Theorem 1:** Given the preceding analysis, proof is a straightforward application of Theorem 3.2 from [23].