Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 56-2 (2023) 1932–1936

Optimisation with Parametric Uncertainty: an ADMM Approach

ZhouDan Pan and **Mark Cannon**

Department of Engineering Science, University of Oxford, Parks Road, Oxford, OX1 3PJ, UK, e-mail: {zhoudan.pan@eng.ox.ac.uk, mark.cannon@eng.ox.ac.uk}

Abstract: To solve convex optimisation problems we rely on iterative algorithms, but when the problem contains parameters that need to be estimated from measurements with noise, another iterative process is needed to perform estimation. In this paper we devise a modified version of the Alternating Direction Method of Multipliers (ADMM) that performs parameter estimation simultaneously with optimisation. Given a convergent parameter estimate, and assuming that the cost function can be expressed in terms of a multi-parametric quadratic program (mp-QP), we prove convergence of the objective values, dual variables and primal residual. Simulation results show that the rate of convergence tracks that of the estimator up to an upper limit that is characterised by convergence rate of ADMM with no parametric uncertainty.

Keywords: Convex optimisation, uncertain systems, distributed control and estimation.

1. INTRODUCTION

Iterative algorithms are widely used to solve convex optimisation problems. Many problems of interest contain parameters that are uncertain and require estimation. For example, consider an optimal control problem in which some parameters of the controlled system are not known exactly but can be estimated from noisy measurements. In this situation it is common to use system identification to estimate unknown plant parameters (e.g. Ljung, 1999), and to wait for the estimated parameters to converge to a sufficiently high level of accuracy before passing the identified model to the decision-making process to compute an optimal control law. In this paper we are concerned with the question of whether we can combine the estimation and optimisation processes in order to accelerate the solution of such problems.

We consider the following optimisation problem:

$$\min_{x,z} \quad f(x) + g(z), \tag{1a}$$

$$\text{subject to} \quad x - z = 0, \tag{1b}$$

where $f, g : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ are convex, closed and proper functions. We assume that some parameters of this problem are unknown, but may be estimated from available measurements. Suppose that $f(x)$ is replaced in (1a) at discrete time instant k by a time-varying function of the form:

$$f^k(x) = \frac{1}{2} x^\intercal Q x + c^\intercal x + \mathcal{I}_{\mathcal{C}^k}(x), \tag{2a}$$

$$\mathcal{C}^k : \quad Ax \le b + D\hat{p}^k. \tag{2b}$$

Here $Q \succeq 0$, $\mathcal{I}_{\mathcal{C}^k}(x)$ is the indicator function of the set $\mathcal{C}^k$ (with $\mathcal{I}_{\mathcal{C}^k}(x) = 0$ for $x \in \mathcal{C}^k$ and $\mathcal{I}_{\mathcal{C}^k}(x) = \infty$ otherwise), and $\hat{p}^k \in \mathcal{P} \subseteq \mathbb{R}^m$ is a random variable representing an estimate of an uncertain parameter in problem (1). We assume that the parameter set $\mathcal{P}$ is compact and that $\hat{p}^k$ can be expressed in terms of a constant but unknown parameter $\bar{p}$ and a random variable δ^k as $\hat{p}^k = \bar{p} + \delta^k$. It is further assumed that δ^k converges almost surely to zero (in particular, the sum over all k of the Euclidean norm, $\|\delta^k\|_2$, is assumed to be finite) so that $\hat{p}^k$ can be considered an asymptotically convergent estimator of $\bar{p}$. We consider the convergence properties of an approach based on the Alternating Direction Method of Multipliers (ADMM Boyd et al., 2010), with $\bar{p}$ approximated by $\hat{p}^k$.

The recent work by Rontsis et al. (2022) solves a semidefinite programming problem using an approximate ADMM solver. However, rather than considering problems with uncertain parameters, the paper uses approximations of the exact ADMM iteration in order to reduce the computational requirements of individual iterations, and this introduces errors into the solution estimates. The paper's analysis shows that convergence is guaranteed provided that the errors introduced in the ADMM iteration are summable across the iterations.

Schulze Darup and Book (2021) investigate using ADMM with a finite number of iterations to solve the optmisation defining a Model Predictive Control (MPC) law. The paper constructs an invariant set containing the target state for the combination of the control problem and the ADMM solver, such that the dynamics of the solver iterations become linear within this set. The paper shows that this simplifies the analysis of the stability of the combined system.

A weighted average consensus algorithm is considered in Esteki et al. (2022) as the basis of a distributed solution for problems involving networked systems with locally time-varying quadratic cost functions. This is motivated by the problem of tracking the optimal solutions of resource allocation problems with time-varying local demand or resource constraints. The paper's approach is able to track the optimal solution in a distributed manner with bounded

Peer review under responsibility of International Federation of Automatic Control.
10.1016/j.ifacol.2023.10.1084

tracking error, with bounds that are dependent on the cost parameters and network topology.

In each of Rontsis et al. (2022), Schulze Darup and Book (2021), and Esteki et al. (2022), the inherent robustness of solver iterations is exploited to derive guarantees of convergence despite bounded uncertainty or random errors in problem data. In this paper we use similar robustness properties to devise a solver for (1). At each iteration, our approach uses the most recent estimate $\hat{p}^k$ to replace the unknown parameter $\bar{p}$ by replacing the uncertain function f with its corresponding estimate f^k. Our approach is applicable more generally to first order methods that can be expressed as Douglas-Rachford splitting (DRS) methods. We show that, if the solver iteration can be expressed in the form of a DRS operator, then, under the assumption of convergent estimation errors, the solution estimate necessarily converges to the solution of (1) corresponding to $\hat{p}^k = \bar{p}$.

1.1 Notation

$\mathbb{R}^n$ denotes the n-dimensional real space. I represents the identity mapping. $Q \succ 0$ and $R \succeq 0$ represent real symmetric positive definite and positive semidefinite matrices. $\mathcal{I}_{\mathcal{C}}(x)$ denotes the indicator function of a closed non-empty set $\mathcal{C}$, so that $\mathcal{I}_{\mathcal{C}}(x) = 0$ for $x \in \mathcal{C}$ and $\mathcal{I}_{\mathcal{C}}(x) = \infty$ otherwise. $\partial F(x)$ indicates the subdifferential of function F evaluated at x. The n-dimensional column vector with all elements equal to 1 is $\mathbf{1}_n$. The truncated normal distribution[1] is denoted $N_{tr}(\mu, \sigma^2, a, b)$. The Euclidean distance between y and $\mathcal{X}$ is denoted $\mathbf{dist}(y, \mathcal{X}) := \inf_{x\in\mathcal{X}} \|x - y\|_2$ and $B_r(q) = \{x \mid \|x - q\|_2 < r\}$ is the open ball of radius r centred on q.

2. PROBLEM FORMULATION

We first consider the solution of the deterministic problem defined by

$$\min_{x,z} \quad \bar{f}(x) + g(z), \tag{3a}$$

$$\text{subject to} \quad x - z = 0, \tag{3b}$$

where $\bar{f}(x) := \frac{1}{2}x^\intercal Qx + c^\intercal x + \mathcal{I}_{\bar{\mathcal{C}}}(x)$ with $\bar{\mathcal{C}} : Ax \le b + D\bar{p}$. This is simply problem (1) with the actual parameter value $\bar{p}$. We construct the augmented Lagrangian for (3):

$$\bar{\mathcal{L}}_\gamma := \bar{f}(x) + g(z) + \lambda^\intercal(x - z) + \frac{1}{2\gamma}\|x - z\|_2^2, \tag{4}$$

in which $\gamma > 0$ is a penalty parameter. By defining the scaled multiplier $u := \gamma\lambda$, the problem (3) can be solved (e.g. Boyd et al., 2010) via the following iterative ADMM updates:

$$x^{k+1} \leftarrow \min_x \mathcal{L}_\gamma(x, z^k, \lambda^k) := \mathbf{prox}_{\gamma\bar{f}}(z^k - u^k), \tag{5a}$$

$$z^{k+1} \leftarrow \min_z \mathcal{L}_\gamma(x^{k+1}, z, \lambda^k) := \mathbf{prox}_{\gamma g}(x^{k+1} + u^k), \tag{5b}$$

$$u^{k+1} \leftarrow u^k + x^{k+1} - z^{k+1}, \tag{5c}$$

where $\mathbf{prox}_{\gamma h}(v) := \arg\min_x \left(h(x) + (1/2\gamma)\|x - v\|_2^2\right)$ denotes the *proximal operator* and k is an iteration counter.

[1] If a random variable x has the normal distribution $N(\mu, \sigma^2)$ and $a < b$, then the distribution of x conditional on $a \le x \le b$ is denoted $N_{tr}(\mu, \sigma^2, a, b)$. We specifically define $x \sim N_{tr}(\mu, \sigma^2, a, a)$ as $\mathbb{P}(x = a) = 1$.

ADMM is well-suited for distributed optimisation when $\bar{f}(x)$ and/or $g(z)$ are separable functions.

By introducing the *reflected proximal operator* $R_{\gamma h} := 2\mathbf{prox}_{\gamma h} - I$, the ADMM iteration (5) can be expressed using *Douglas-Rachford Splitting* (DRS) in terms of an operator $\bar{T} : \mathbb{R}^n \to \mathbb{R}^n$ defined by

$$v^{k+1} = \bar{T}v^k := \left(\frac{1}{2}I + \frac{1}{2}R_{\gamma\bar{f}}R_{\gamma g}\right)v^k, \tag{6}$$

where $v^k := x^k + u^{k-1}$, and z, x, u in (5) can be obtained given the value of v (see Rontsis et al. (2022) and Giselsson et al. (2016), Appendix B) as:

$$x^k = \mathbf{prox}_{\gamma\bar{f}} R_{\gamma g} v^{k-1}, \tag{7a}$$

$$z^k = \mathbf{prox}_{\gamma g} v^k, \tag{7b}$$

$$u^k = \left(I - \mathbf{prox}_{\gamma g}\right) v^k. \tag{7c}$$

Note that, as a result of DRS, $\bar{T}$ is a *firmly non-expansive* operator (see e.g. Bauschke et al., 2012) that satisfies, $\forall v, w \in \mathbb{R}^n$,

$$\|\bar{T}v - \bar{T}w\|_2^2 + \|(I - \bar{T})v - (I - \bar{T})w\|_2^2 \le \|v - w\|_2^2. \tag{8}$$

3. ADMM WITH PARAMETRIC UNCERTAINTY

We suppose $\hat{p}^k$ in (2) is a real-time estimate of $\bar{p}$ in (3), with

$$\hat{p}^k = \bar{p} + \delta^k \in \mathcal{P}, \tag{9}$$

where $\mathcal{P}$ is assumed to be compact. Convergent estimators typically ensure that $\|\delta^k\|_2 \to 0$ almost surely (a.s.) as $k \to \infty$, and we therefore expect to obtain an accurate estimate of $\bar{p}$ by running the estimator for a sufficiently long time. By subsequently passing this estimate to the ADMM solver and iterating for another sufficiently large number of iterations we expect to obtain an approximate solution to problem (3). In this section we consider how to combine the two processes to reliably accelerate the overall solution process for real-world applications with estimated problem data.

By integrating $\hat{p}^k$ into the ADMM iteration (5) we propose the following ADMM algorithm with parametric uncertainty (ADMM-PU):

$$x^{k+1} \leftarrow \mathbf{prox}_{\gamma f^k}(z^k - u^k), \tag{10a}$$

$$z^{k+1} \leftarrow \mathbf{prox}_{\gamma g}(x^{k+1} + u^k), \tag{10b}$$

$$u^{k+1} \leftarrow u^k + x^{k+1} - z^{k+1}, \tag{10c}$$

in which $f^k(x) := f(x)|_{p^k=\hat{p}^k}$. Note that we assume for convenience that the solver and estimator share a common iteration index k; more generally $\hat{p}^k$ denotes the most recently available estimate.

Similar to (6)-(7), the ADMM-PU iteration (10) can be expressed as a time-varying DRS operator $T^k : \mathbb{R}^n \to \mathbb{R}^n$:

$$v^{k+1} = T^k v^k := \left(\frac{1}{2}I + \frac{1}{2}R_{\gamma f^k}R_{\gamma g}\right)v^k, \tag{11}$$

where $v^k := x^k + u^{k-1}$, and x, z, u can be obtained from:

$$x^k = \mathbf{prox}_{\gamma f^k} R_{\gamma g} v^{k-1}, \tag{12a}$$

$$z^k = \mathbf{prox}_{\gamma g} v^k, \tag{12b}$$

$$u^k = \left(I - \mathbf{prox}_{\gamma g}\right) v^k. \tag{12c}$$

The following lemma shows that T^k can be expressed in terms of a bounded perturbation of the operator $\bar{T}$ defined in (6).

Lemma 1. The iteration (10) can be represented as
$$v^{k+1} = \bar{T}v^k + E^k\delta^k,$$
where $E^k \in \mathbb{R}^{n\times m}$ satisfies $\|E^k\|_2 \le \bar{e}$ for some finite $\bar{e}$.

Proof. Since (10) can be expressed as (11), we have
$$\begin{aligned} v^{k+1} = T^k v^k &= \frac{1}{2}v^k + \frac{1}{2}R_{\gamma f^k}R_{\gamma g}v^k \\ &= \frac{1}{2}v^k + \frac{1}{2}(2\mathbf{prox}_{\gamma f^k} - I)R_{\gamma g}v^k. \end{aligned} \tag{13}$$
For all $y \in \mathbb{R}^n$, $\mathbf{prox}_{\gamma f^k}(y)$ solves the following problem
$$\min_x \quad \frac{1}{2}x^\intercal Qx + c^\intercal x + \frac{1}{2\gamma}\|x-y\|_2^2, \tag{14a}$$
$$\text{subject to} \quad Ax \le b + D\hat{p}^k = b + D\bar{p} + D\delta,\ \delta = \delta^k. \tag{14b}$$
Define a matrix H, vector w and linear functional s as follows,
$$H := Q + \gamma^{-1}I \quad (Q \succeq 0 \text{ and } \gamma > 0 \Rightarrow\ H \succ 0), \tag{15a}$$
$$w := x + H^{-1}(c - \gamma^{-1}y), \tag{15b}$$
$$s(y) := AH^{-1}(c - \gamma^{-1}y) + b + D\bar{p}. \tag{15c}$$
Then, by completing the square, solving (14) is equivalent to solving
$$\min_w \quad \frac{1}{2}w^\intercal Hw, \tag{16a}$$
$$\text{subject to} \quad Aw \le s(y) + D\delta,\ \delta = \delta^k, \tag{16b}$$
and obtaining x from (15b). Since (16) takes the standard form of a right-hand side *multi-parametric quadratic program* (mp-QP), the optimiser $w^*(\delta) : \mathbb{R}^m \to \mathbb{R}^n$ is continuous and piecewise affine, and can be expressed in the following form (see Bemporad et al., 2002):
$$w^*(\delta) = H^{-1}\tilde{A}^\intercal(\tilde{A}H^{-1}\tilde{A}^\intercal)^{-1}(\tilde{s}(y) + \tilde{D}\delta). \tag{17}$$
Here $\tilde{A}$, $\tilde{s}(y)$ and $\tilde{D}$ correspond to the set of active constraints that depends on δ, and we assume the rows of $\tilde{A}$ are linearly independent. Since $w^*(\delta)$ is piecewise affine, $w^*(\delta^k)$ can be expressed as:
$$w^*(\delta^k) = w^*(0) + E^k\delta^k, \tag{18}$$
where
$$w^*(0) = \mathbf{prox}_{\gamma\bar{f}}(y) + H^{-1}(c - \gamma^{-1}y), \tag{19}$$
and E^k can be determined by considering the value of $\partial w^*(\delta)$ along the line segment L from $\delta = 0$ to $\delta = \delta^k$. Without loss of generality we impose $E^k = 0$ when $\delta^k = 0$. For $\delta^k \neq 0$, the mean value theorem (Rockafeller and Wets, 1997) implies that $E^k = \partial w^*(\bar{\delta})$, for some $\bar{\delta} \in L$. Since $\partial w^*(\delta) \ni H^{-1}\tilde{A}^\intercal(\tilde{A}H^{-1}\tilde{A}^\intercal)^{-1}\tilde{D}$ is independent of y we have
$$\|E^k\|_2 \le \max_{j\in J}\|H^{-1}\tilde{A}(j)^\intercal(\tilde{A}(j)H^{-1}\tilde{A}(j)^\intercal)^{-1}\tilde{D}(j)\|_2 := \bar{e}$$
where index j indicates a given set of active constraints, and J is the set of indices of all possible active constraint sets for problem (16). Note that $\bar{e}$ is necessarily finite since H in (15) is positive definite.

Combining (15b) and (18)-(19) we obtain
$$\begin{aligned} \mathbf{prox}_{\gamma f^k}(y) &= w^*(0) - H^{-1}(c - \gamma^{-1}y) + E^k\delta^k \\ &= \mathbf{prox}_{\gamma\bar{f}}(y) + E^k\delta^k. \end{aligned} \tag{20}$$
Substituting (20) into (13) then gives the result
$$T^k v^k = \frac{1}{2}v^k + \frac{1}{2}R_{\gamma\bar{f}}R_{\gamma g}v^k + E^k\delta^k = \bar{T}v^k + E^k\delta^k, \tag{21}$$
where $\|E^k\|_2 \le \bar{e}$. □

Our assumptions that the feasible set of problem (2) is non-empty and compact, and that the error in the estimate of $\bar{p}$ in problem (3) is bounded and convergent, can be stated as follows.

Assumption 1. For all k we require that:

(a) $\mathcal{C}^k$ is non-empty and satisfies, for finite r,
$$\mathcal{C}^k = \{x \mid Ax \le b + D\hat{p}^k\} \subset B_r(0),$$
(b) $\delta^k \in B_\rho(0)$ and $\sum_{i=0}^{k}\|\delta^i\|_2 \le \Delta$ a.s., for finite ρ and Δ.

Let $\mathbf{Fix}(\bar{T})$ denote the fixed point set of $\bar{T}$ that corresponds to the set of saddle points of (4). Before proving convergence of ADMM-PU, we first show that the iteration (11) ensures $\|\bar{T}v^k - v^*\|_2$ is finite for all k, where v^* is any point in $\mathbf{Fix}(\bar{T})$ such that $\|v^0 - v^*\|_2$ is finite.

Lemma 2. Under Assumption 1 the iteration (11) satisfies, for all k,
$$\|\bar{T}v^k - v^*\|_2 \le \|v^0 - v^*\|_2 + \bar{e}\Delta, \tag{22}$$
where $v^* \in \mathbf{Fix}(\bar{T})$.

Proof. From Lemma 1 we have
$$\begin{aligned} \|v^{k+1} - v^*\|_2 &= \|\bar{T}v^k + E^k\delta^k - v^*\|_2 \\ &\le \|\bar{T}v^k - v^*\|_2 + \|E^k\delta^k\|_2 \\ &\le \|\bar{T}v^k - v^*\|_2 + \bar{e}\|\delta^k\|_2. \end{aligned}$$
But (8) implies that $\bar{T}$ is contractive, since $v = v^k$ and $w = v^*$ in (8) gives $\|\bar{T}v^k - v^*\|_2 \le \|v^k - v^*\|_2$, and hence
$$\|v^{k+1} - v^*\|_2 \le \|v^k - v^*\|_2 + \bar{e}\|\delta^k\|_2.$$
Summing both sides of this inequality over $0 \le k \le N$, we obtain
$$\|v^{N+1} - v^*\|_2 \le \|v^0 - v^*\|_2 + \bar{e}\sum_{k=0}^{N}\|\delta^k\|_2.$$
From Assumption 1(b), it follows that, for all k,
$$\|v^k - v^*\|_2 \le \|v^0 - v^*\|_2 + \bar{e}\Delta,$$
and (22) then follows from the contractivity of $\bar{T}$, which implies $\|\bar{T}v^k - v^*\|_2 \le \|v^k - v^*\|_2$.

Theorem 3. Under Assumption 1, the ADMM-PU iteration (10) converges as $k \to \infty$, with:

(i) $\|\text{obj}^k - \text{obj}^*\|_2 \to 0$ a.s.
(ii) $\|\lambda^k - \lambda^*\|_2 \to 0$ a.s.
(iii) $\|x^k - z^k\|_2 \to 0$ a.s.

where obj^* and λ^* are respectively the optimal values of the objective and dual variable for problem (3), and $\text{obj}^k := \text{obj}(x^k, z^k) := \frac{1}{2}{x^k}^\intercal Qx^k + c^\intercal x^k + g(z^k)$.

Remark 4. The ADMM iteration (5), in which $\bar{p}$ is known, provides deterministic guarantees of convergence of obj^k, λ^k, x^k and z^k analogous to (i)-(iii) (e.g. Boyd et al., 2010)).

Remark 5. Assumption 1(b) requires that $\|\delta^k\|_2$ converges sufficiently rapidly to zero as $k \to \infty$. Since $\|\delta^k\|_2 \le \|\delta^k\|_1$, this convergence assumption is satisfied by any l_1-stable estimator (such as an exponentially stable Luenberger observer, for example) driven by a finite l_1-norm noise sequence (e.g. Vidyasagar, 1992, Sec. 6.7).

Proof. The proof of Theorem 3 is given in two parts.

Part I (to prove that $\|(I-\bar{T})v^k\|_2^2 \to 0$ a.s.): Let v^* be any point in $\mathbf{Fix}(\bar{T})$ such that $\|v^0 - v^*\|_2$ is finite. Then the firmly non-expansive property of $\bar{T}$ implies (e.g. by setting $v = v^k$ and $w = v^*$ in (8)) that

$$\|\bar{T}v^k - v^*\|_2^2 + \|(I-\bar{T})v^k\|_2^2 \le \|v^k - v^*\|_2^2. \tag{23}$$

But Lemma 1 implies

$$\begin{aligned}&\|v^{k+1} - v^*\|_2^2 - \|\bar{T}v^k - v^*\|_2^2 \\ &\quad = \|E^k\delta^k\|_2^2 + 2(E^k\delta^k)^\intercal(\bar{T}v^k - v^*) \\ &\quad \le \|E^k\delta^k\|_2\big(\|E^k\delta^k\|_2 + 2\|\bar{T}v^k - v^*\|_2\big).\end{aligned}$$

Assumption 1(b) and (22) therefore imply

$$\begin{aligned}&\|v^{k+1} - v^*\|_2^2 - \|\bar{T}v^k - v^*\|_2^2 \\ &\quad \le \bar{e}\|\delta^k\|_2\big(\bar{e}\rho + 2\|v^0 - v^*\|_2 + 2\bar{e}\Delta\big) \\ &\quad = \kappa\,\|\delta^k\|_2,\end{aligned} \tag{24}$$

where

$$\kappa = \bar{e}^2(\rho + 2\Delta) + 2\bar{e}\|v^0 - v^*\|_2.$$

Combining the inequality (24)) with (23) yields

$$\|v^{k+1} - v^*\|_2^2 + \|(I-\bar{T})v^k\|_2^2 \le \|v^k - v^*\|_2^2 + \kappa\|\delta^k\|_2.$$

Summing both sides of this inequality for $0 \le k \le N$ yields

$$\begin{aligned}\sum_{k=0}^{N}\|(I-\bar{T})v^k\|_2^2 &\le \|v^0 - v^*\|_2^2 + \kappa\sum_{k=0}^{N}\|\delta^k\|_2 \\ &\le \|v^0 - v^*\|_2^2 + \kappa\Delta\end{aligned}$$

for all N. Since the right hand side is finite for all N, this implies that $\|(I-\bar{T})v^k\|_2^2 \to 0$ a.s. as $k \to \infty$.

Part II (to prove properties (i)-(iii) in Theorem 3): Since the reference problem (3) satisfies the *Linear Independent Constraint Qualification* (LICQ), the optimal dual variable λ^* in (4) is unique (see Wachsmuth, 2013)). Additionally, since Assumption 1(a) implies that the feasible set of the convex optimisation problem (3) is compact, the set of all optimal solutions $x^* \in \mathcal{X}$ is compact and convex. Hence $\mathbf{Fix}(\bar{T}) = \{v^* \,|\, \forall x^* \in \mathcal{X}, v^* = x^* + \gamma\lambda^*\}$ is convex and compact. Therefore, for any given $\epsilon > 0$, choose $r \in (0, \epsilon]$ and consider the following compact set,

$$\Omega_r := \{v \in \mathbb{R}^n \,|\, \mathbf{dist}(v, \mathbf{Fix}(\bar{T})) \le r\}. \tag{25}$$

Let $\alpha = \min_{\mathbf{dist}(v,\mathbf{Fix}(\bar{T}))=r}\|(I-\bar{T})v\|_2^2$. Choose $\beta \in (0, \alpha)$ and let

$$\Omega_\beta := \{v \in \Omega_r \,|\, \|(I-\bar{T})v\|_2^2 \le \beta\}. \tag{26}$$

Then, since Ω_β is contained in the interior of Ω_r, we have $\mathbf{dist}(v, \mathbf{Fix}(\bar{T})) < \epsilon$ for all $v \in \Omega_\beta$. Furthermore, Part I showed that $\|(I-\bar{T})v^k\|_2^2 \to 0$ a.s., and it follows that

$$\mathbf{dist}(v^k, \mathbf{Fix}(\bar{T})) \to 0 \text{ a.s.} \tag{27}$$

Moreover, Assumption 1(b) implies $\|\delta^k\|_2 \to 0$ a.s. so (20) and Lemma 1 imply, for all $v \in \mathbb{R}^n$,

$$\|(\mathbf{prox}_{\gamma f^k} - \mathbf{prox}_{\gamma\bar{f}})v\|_2 \to 0 \text{ a.s.} \tag{28a}$$

$$\|(T^k - \bar{T})v\|_2 \to 0 \text{ a.s.} \tag{28b}$$

By substituting (28) into (12) we therefore have

$$\|x^k - \bar{x}(v^{k-1})\|_2 \to 0 \text{ a.s.}, \quad \bar{x}(v^{k-1}) := \mathbf{prox}_{\gamma\bar{f}}R_{\gamma g}v^{k-1}.$$

To complete the proof, we use the continuity of the objective function to infer that

$$\|\text{obj}(x^k, z^k) - \text{obj}(\bar{x}(v^{k-1}), z^k)\|_2 \to 0 \text{ a.s.}$$

and, by (27),

$$\|\text{obj}(\bar{x}(v^{k-1}), z^k) - \text{obj}^*\|_2 \to 0 \text{ a.s.}$$

and hence that $\|\text{obj}^k - \text{obj}^*\|_2 \to 0$ a.s. By an analogous argument we can also conclude that $\|\lambda^k - \lambda^*\|_2 \to 0$ and $\|x^k - z^k\|_2 \to 0$ a.s. □

4. NUMERICAL STUDY

This section investigates the convergence of the proposed algorithm using numerical simulations. The example we consider is the following resource allocation problem:

$$\min_{x,z} \quad (z - s^m)^\intercal Q(z - s^m) + c^\intercal z \tag{29a}$$

$$\text{subject to} \quad x = z \tag{29b}$$

$$\mathbf{1}_n^\intercal x = p \tag{29c}$$

$$s^l \le z \le s^u \tag{29d}$$

where $x, z, s^l, s^u, s^m \in \mathbb{R}^n$, $Q := \mathbf{diag}(q_1, q_2, \ldots, q_n) \succeq 0$, and $p \in \mathbb{R}$. This problem can be viewed as n decentralised energy suppliers collaborating in order to match an unknown demand $\bar{p}$. The parameters of the problem are generated as follows:

$$\forall i : \begin{cases} q_i \sim U(0,2),\ c_i \sim U(400,600), \\ s_i^m \sim U(80,120),\ s_i^r \sim U(20,30), \\ s_i^l = s_i^m - s_i^r,\ s_i^u = s_i^m + s_i^r, \end{cases} \tag{30a}$$

$$\bar{p} \sim U(\mathbf{1}_n^\intercal(s^m - 0.5s^r), \mathbf{1}_n^\intercal(s^m + 0.5s^r)), \tag{30b}$$

with the modification that $q_1 = 0$ so that problem (29) is not strongly convex. We choose $n = 10$, and set $\gamma = 0.1$ as the penalty term for ADMM iterations.

The following estimator models are considered for $\hat{p}^k$:

Model 1: $\hat{p}^k \sim N_{tr}(\bar{p}, (10e^{-\frac{k}{5}})^2, d^l, d^u)$,

Model 2: $\hat{p}^k \sim N_{tr}(\bar{p}, (10e^{-\frac{k}{10}})^2, d^l, d^u)$,

Model 3: $\hat{p}^k = \dfrac{1}{k}\sum_{i=1}^{k} p^i_{\text{meas}},\ p^i_{\text{meas}} \sim N_{tr}(\bar{p}, 1^2, d^l, d^u)$,

where $d^l = \bar{p} - \frac{1}{2}\mathbf{1}_n^\intercal s^r$, $d^u = \bar{p} + \frac{1}{2}\mathbf{1}_n^\intercal s^r$. Models 1 and 2 can be interpreted as estimators that converge in mean with high (Model 1) and low (Model 2) convergence rates. Since the probability distribution of $\hat{p}^k$ (and hence that of δ^k) has constant support for all k, the almost sure convergence requirement of Assumption 1(b) does not hold. However, bounds on δ^k corresponding to any confidence level less than 1 for Models 1 and 2 are exponentially convergent and thus have finite l^1-norm. Model 3 estimates $\bar{p}$ by taking the running average of the i.i.d. measurements represented by p^k_{meas}. As in the case of Models 1 and 2, Model 3 does not satisfy Assumption 1(b) (almost sure convergence), and moreover the bounds on δ^k corresponding to a confidence level less than 1 converge at the rate of $1/\sqrt{k}$ in this case, and therefore do not have finite l^1-norm.

We sample (30) for one instance and pass this instance to the proposed ADMM-PU algorithm (10) with each of the estimator models. The evolution of the fractional error in the objective value, $|\text{obj}^k - \text{obj}^*|/\text{obj}^*$, representing the residual error after k iterations, is shown in Fig.1. For benchmarking purposes, we compute (using any capable solver) at each time step k the solution of the deterministic problem defined by (2) with the most recent estimate $\hat{p}^k$, and this is denoted as $\text{opt}(\hat{p}^k)$ in Fig. 1. We also run the standard ADMM iteration (5) with the true value of $\bar{p}$ to solve problem (3) directly, and the evolution of these iterations is shown in Fig. 1 as ADMM-ref.

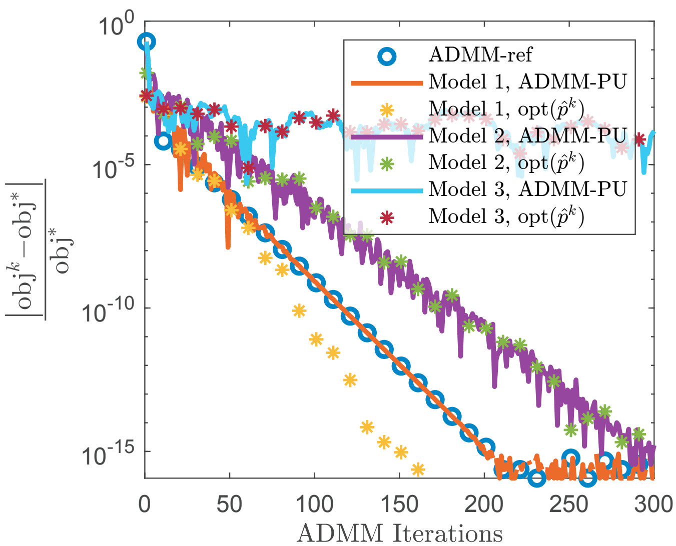

Fig. 1. Simulation results

From the residual errors obtained with the Model 1 estimator (Fig. 1), it can be seen that ADMM-PU converges at the same rate as ADMM-ref, which is slower than opt($\hat{p}^k$). This is to be expected, because with the Model 1 estimator, $\hat{p}^k$ converges faster than ADMM-ref, while the convergence rate of ADMM-ref is necessarily an upper bound for the convergence rate of ADMM-PU. With the Model 2 estimator, $\hat{p}^k$ converges more slowly than ADMM-ref, and the proposed ADMM-PU algorithm is therefore able to track opt($\hat{p}^k$) closely. With the Model 3 estimator, the estimation error δ^k converges more slowly than for Models 1 and 2. However, Fig. 1 shows that ADMM-PU is sufficiently robust to track opt($\hat{p}^k$) in this case.

5. CONCLUSIONS

This paper proposes a variant of ADMM which combines simultaneous iterations for optimisation and parameter estimation. The proposed approach guarantees convergence almost surely, provided that the sub-problem solved at each ADMM iteration takes the form of a mp-QP and the estimation error converges almost surely. Simulation results demonstrate that the proposed algorithm tracks the solution of the deterministic problem defined in terms of the most recent parameter estimate, provided the estimator converges more slowly than the standard ADMM iteration with no parameter uncertainty. For cases in which the estimator converges faster than this, the proposed algorithm converges at its maximum rate, which is the same rate as the standard ADMM iteration with no parameter uncertainty.

There are various directions to extend this work. Lemma 1 implies that the *proximal operator* of the time varying objective function (2) (i.e. $\mathbf{prox}_{\gamma f^k}(\cdot)$) can be expressed in terms of the proximal operator of the original objective of problem (1) and a remainder term that is bounded by a linear function of the estimator error. This property is used in Theorem 3 to prove convergence of the proposed algorithm based on ADMM. However, the proximal operator is the *resolvent* of the subdifferential operator and appears in various splitting methods other than DRS (see Ryu and Boyd, 2016). This analysis approach can therefore be applied more generally to first order methods other than ADMM.

Apart from parametric uncertainty, the parameter estimation error may also arise due communication delays in a distributed optimisation setting. Peng et al. (2016) put forward an algorithmic framework for asynchronous iteration updates, and we are therefore able investigate the convergence behavior when delays and parametric uncertainty coexist through an extension of the current work. We also plan to investigate how the algorithm works in a real-world context, e.g. applications in power systems with uncertainty or online collaboration of robots.

REFERENCES

Bauschke, H.H., Moffat, S.M., and Wang, X. (2012). Firmly Nonexpansive Mappings and Maximally Monotone Operators: Correspondence and Duality. *Set-Valued and Variational Analysis*, 20(1), 131–153. doi: 10.1007/s11228-011-0187-7.

Bemporad, A., Morari, M., Dua, V., and Pistikopoulos, E.N. (2002). The explicit linear quadratic regulator for constrained systems. *Automatica*, 38(1), 3–20. doi: 10.1016/S0005-1098(01)00174-1.

Boyd, S., Parikh, N., Chu, E., Peleato, B., and Eckstein, J. (2010). Distributed optimization and statistical learning via the alternating direction method of multipliers. doi: 10.1561/2200000016.

Esteki, A.s., Kia, S.S., and Member, S. (2022). Distributed Optimal Resource Allocation with Time-Varying Quadratic Cost Functions and Resources over Switching Agents. In *2022 European Control Conference (ECC)*, 441–446.

Giselsson, P., Fält, M., and Boyd, S. (2016). Line Search for Averaged Operator Iteration. *ArXiv*. URL `http://arxiv.org/abs/1603.06772`.

Ljung, L. (1999). *System Identification – Theory for the User.* Prentice Hall, Upper Saddle River, NJ.

Peng, Z., Xu, Y., Yan, M., and Yin, W. (2016). ARock: An Algorithmic Framework for Asynchronous Parallel Coordinate Updates. *SIAM Journal on Scientific Computing*, 38(5), A2851–A2879. doi:10.1137/15M1024950. URL `http://epubs.siam.org/doi/10.1137/15M1024950`.

Rockafeller, R.T. and Wets, R.J.B. (1997). *Convex analysis.* Princeton, NJ: Princeton University Press.

Rontsis, N., Goulart, P., and Nakatsukasa, Y. (2022). Efficient Semidefinite Programming with Approximate ADMM. *Journal of Optimization Theory and Applications*, 192(1), 292–320. doi:10.1007/s10957-021-01971-3.

Ryu, E.K. and Boyd, S. (2016). a Primer on Monotone Operator Methods Survey. *Appl. Comput. Math*, 3–43.

Schulze Darup, M. and Book, G. (2021). On Closed-Loop Dynamics of ADMM-Based MPC. In *Lecture Notes in Control and Information Sciences*, volume 485, 107–134. Springer Science and Business Media Deutschland GmbH. doi:10.1007/978-3-030-63281-6_5.

Vidyasagar, M. (1992). *Nonlinear systems analysis.* Prentice Hall, Upper Saddle River, NJ, 2nd edition.

Wachsmuth, G. (2013). On LICQ and the uniqueness of Lagrange multipliers. *Operations Research Letters*, 41(1), 78–80. doi:10.1016/j.orl.2012.11.009.