

7-5-2013

# Lagrangian relaxation based decomposition for well scheduling in shale-gas systems

Brage Rugstad Knudsen  
*Norwegian University of Science and Technology*

Ignacio E. Grossmann  
*Carnegie Mellon University, grossmann@cmu.edu*

Bjarne Foss  
*Norwegian University of Science and Technology*

Andrew R. Conn  
*IBM*

Follow this and additional works at: <http://repository.cmu.edu/cheme>

 Part of the [Chemical Engineering Commons](#)

---

## Published In

Computers and Chemical Engineering, 63, 234-249.

This Article is brought to you for free and open access by the Carnegie Institute of Technology at Research Showcase @ CMU. It has been accepted for inclusion in Department of Chemical Engineering by an authorized administrator of Research Showcase @ CMU. For more information, please contact [research-showcase@andrew.cmu.edu](mailto:research-showcase@andrew.cmu.edu).

# Lagrangian Relaxation Based Decomposition for Well Scheduling in Shale-gas Systems

Brage Rugstad Knudsen<sup>a,\*</sup>, Ignacio E. Grossmann<sup>b</sup>, Bjarne Foss<sup>a</sup>, Andrew R. Conn<sup>c</sup>

<sup>a</sup>*Department of Engineering Cybernetics, Norwegian University of Science and Technology, Trondheim, Norway.*

<sup>b</sup>*Department of Chemical Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, USA.*

<sup>c</sup>*IBM T. J. Watson Research Center, Yorktown Heights, NY, USA.*

---

## Abstract

Suppressing the effects of liquid loading is a key issue for efficient utilization of mid and late-life wells in shale-gas systems. This state of the wells can be prevented by performing short shut-ins when the gas rate falls below the minimum rate needed to avoid liquid loading. In this paper, we present a Lagrangian relaxation based scheme for shut-in scheduling of distributed shale multi-well systems. The scheme optimizes shut-in times and a reference rate for each multi-well pad, such that the total produced rate tracks a given short-term gas demand for the field. By using simple, frequency-tuned well proxy models, we obtain a compact mixed integer formulation which by Lagrangian relaxation renders a decomposable structure. A set of computational tests demonstrates the merits of the proposed scheme. This study indicates that the method is capable of solving large field-wide scheduling problems by producing good solutions in reasonable computation times.

## Keywords:

Shale-gas production, Lagrangian relaxation, mixed-integer programming, parameter estimation.

---

## 1. Introduction

The use of shale-gas as an energy resource has increased extensively over the last decade [1]. Even though shale-gas recovery has primarily been a U.S. driven industry, there is currently an increasing exploration of shale and other tight formation resources both in the Middle East, in the North Africa, in China as well as in Europe [2]. The long horizontal wells and stimulation with hydraulic fracturing needed to obtain profitable recovery rates from the very tight rock formations, is however very costly. The last years' low gas prices have therefore made many shale-gas fields barely economical. This clearly increases the need for efficient planning, exploration and production techniques in shale-gas recovery.

Modern shale-gas developments are characterized by an increasing use of multi-well pads for shearing of surface infrastructure, both in the production from the wells and during the completion [3]. Multi-well pads consist of several wells drilled at a single location, hence reducing

---

\*Corresponding author

Email address: [brage.knudsen@itk.ntnu.no](mailto:brage.knudsen@itk.ntnu.no) (Brage Rugstad Knudsen)

Preprint submitted to Elsevier

July 5, 2013

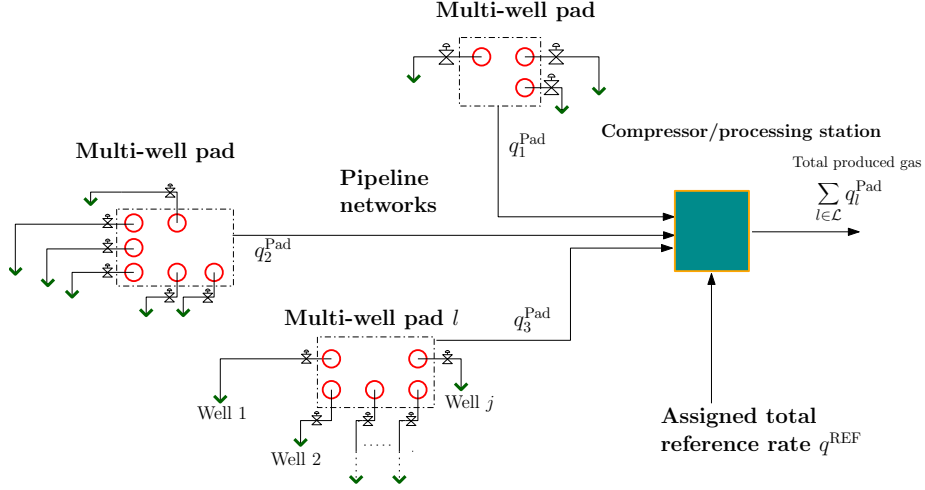


Figure 1: Illustration of three pad shale-gas system.

the foot-print at the well locations by requiring less space for surface equipment, while at the same time encompassing a substantial part of the valuable shale formation. The pads are normally spread over a wide geographical area, with interconnecting pipeline networks transporting the gas to a compressor or processing station. This infrastructure is illustrated in Figure 1. The gas-rate production profile from shale-gas wells are characterized by an initial high peak-rate or plateau rate if choking is performed, with a early steep decline followed by low pseudo steady-state rates [4; 5]. Shale-gas wells may produce at this pseudo steady-state for some hundred days or even for many years. All shale and tight formation gas wells will, however, eventually reach the state where the wells' pressure is insufficient to lift co-produced water to the surface, causing liquid accumulation in the vertical wellbore [6]. This state is known as liquid loading, and observed by very erratic and unstable rate, sharp drops in the decline curve [7; 8] or semi-stabilization of the production rate at significantly lower levels [6; 9], so-called meta-stable rates. Liquid loading severely deteriorates the performance of gas wells, and requires some remedial operational procedures. Whitson et al. [6] showed that liquid loading can be eliminated by performing short regular shut-ins when the producing gas rate reaches the minimum rate required for continuous removal of liquids in the wellbore. This scheme is illustrated for a single well in Figure 2. Whitson et al. [6] also demonstrated that a cyclic scheme leads to minimal loss in cumulative production for low permeability wells when comparing to continuously-unloaded ("ideal" and non-realistic) production, while a significant improvement was obtained when comparing the cyclic with production at meta-stable rates.

The cyclic shut-in scheme introduced by [6] was extended to multi-well systems and time-varying shut-ins in [10]. The problem of optimizing shut-in times for shale multi-well pads was formulated as a mixed integer linear program (MILP), using simplified well and reservoir proxy models [11]. For field-wide multi-pad systems, however, the dimensionality of the MILP becomes computationally prohibitive. The distributed shale-gas multi-pad system shown in Figure 1 renders a network-type structure with limited coupling between the pads. For these types of systems, the use of a Lagrangian relaxation method provides an efficient solution approach by decomposing the original problem into smaller subproblems, see e.g. [12; 13]. Still, any effi-

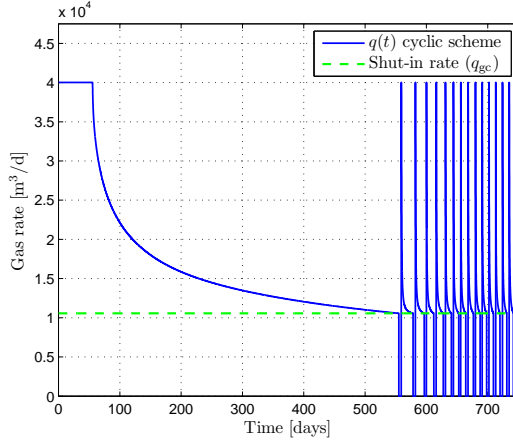


Figure 2: Illustration of a cyclic shut-in scheme, showing the simulated gas rate for a multi-fractured horizontal shale-gas well using a *fixed* shut-in time.

cient implementation of a Lagrangian relaxation requires a stable and reliable method for finding optimal Lagrangian multipliers, as well as a heuristic for finding primal feasible solutions based on the solution from solving the Lagrangian and the dual. Furthermore, for the type of mixed integer dynamic system considered in this paper, an efficient optimization scheme relies on using sufficiently accurate and numerically efficient proxy models, together with a good problem formulation such that the subproblems can be solved in reasonable computation time. The goal of this paper is to address these issues in the context of shut-in based production planning for the shale-gas multi-pad system shown in Figure 1.

The remainder of this paper is organized as follows. Section 2 presents a review of the shale well and reservoir proxy model. In Section 3, we present a parameter estimation problem for tuning of the proxy model, together with a simple scheme for updating the model parameters. Section 4 presents the formulation of fullspace optimization problem, while the Lagrangian relaxation scheme is presented in Section 5. Subsequently, results from computational testing of the scheme will be presented in Section 6. A discussion on the applicability and the numerical efficiency of the proposed scheme is given in Section 7, with concluding remarks in Section 8 ending the paper.

## 2. Shale well and reservoir proxy modeling

Hydraulically fractured shale-gas reservoirs are conventionally modeled either as variations of the dual-porosity idealization of naturally fractured reservoirs [14], as discrete fracture models [15], or as fully discretized single-porosity dual-permeability models [16]. These modeling techniques have different merits and drawbacks in terms of accuracy, detail level and numerical efficiency, while they are all constructed with the incentive of both capturing highly dynamic short-term effects as well as predicting long-term recovery rates. In contrast to these detailed, numerically demanding modeling schemes, is the widespread use of simple semi-analytical or empirical models for gas-rate forecasting [7; 17; 18]. These models, however, assume continuous production at a constant bottomhole pressure, and are therefore not suited for optimization of a cyclic production scheme.

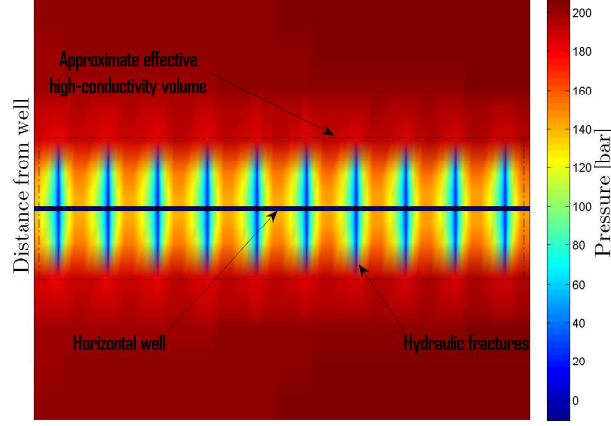


Figure 3: The topside view of the grid pressure after 450 days of constant production from simulation of a multi-fracture shale-gas reservoir model.

Both of the aforementioned groups of shale-gas modeling schemes are efficient for their purposes, but lack applicability with respect to the goal of efficiently optimizing shut-ins for large multi-well systems. A simplified well and reservoir proxy model was therefore derived in [11], with modifications and extensions in [10]. Consider Figure 3 showing the grid pressure after 450 days of constant production in a simulated multi-fracture shale-gas reservoir model. The pressure depletion around and between the fractures will over time "homogenize" the stimulated reservoir volume encompassing the fracture system, leading to a determination of average properties and pressures in this region of the reservoir. During shut-ins, gas will flow from the outer low permeability region with high pressure (red in Figure 3) and recharge the effective low-pressure fracture system volume with gas. The reservoir proxy model is based on approximating the effective low-pressure high-conductivity volume marked out in Figure 3 as *one*, compact volume. By using a radial composite model as shown in Figure 4, we construct the proxy as three-region model with an intermediate crushed rock-type volume and a low permeable outer region representing the tight shale rock. The effective high-conductivity volume in Figure 3 will vary and clearly increase at late times of the well.

Assuming Darcys law and single phase gas together with the law of mass conservation and the equation of state for real gases, the nonlinear partial differential equation (PDE) for the gas flow in the radial composite proxy model in Figure 4 is given by

$$\phi \frac{p}{Z(p)} c(p) \frac{\partial p}{\partial t} = \frac{1}{r} \frac{\partial}{\partial r} \left( k(r) \frac{p}{\mu(p)Z(p)} r \frac{\partial p}{\partial r} \right), \quad (1)$$

where  $p$  is the pressure,  $k(r)$  is a radially dependent permeability,  $\mu(p)$  is the gas viscosity,  $Z(p)$  is the gas compressibility factor and  $c(p)$  is the total compressibility. The pressure dependency of  $\mu(p)$  and  $Z(p)$  can be efficiently handled by the variable transformation known as *pseudopressure* [19],

$$m(p) = 2 \int_0^p \frac{p'}{\mu(p')Z(p')} dp'. \quad (2)$$

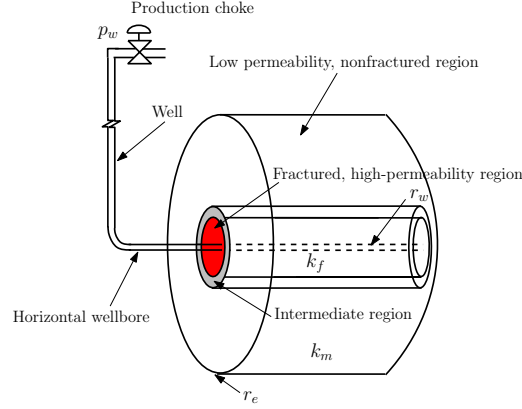


Figure 4: Illustration of proxy model

By using Neumann boundary conditions with a producing well (sink) at the center  $r_w$ , no-flow conditions at the outer boundary  $r_e$ , and a given initial reservoir pressure, we obtain the following initial boundary-value problem (IBVP) in terms of the reservoir pseudopressure  $m(p)$  [11].

$$\phi\mu c \frac{\partial m}{\partial t} = \frac{1}{r} \frac{\partial}{\partial r} \left( k(r)r \frac{\partial m}{\partial r} \right), \quad (3a)$$

$$r \frac{\partial m}{\partial r} \Big|_{r_w} = q \frac{T p_{sc}}{T_{sc} \pi h k}, \quad (3b)$$

$$\frac{\partial m}{\partial r} \Big|_{r_e} = 0, \quad (3c)$$

$$m(r, 0) = m_{\text{init}}. \quad (3d)$$

where  $q$  is the gas rate,  $T$  is reservoir temperature and  $h$  is the horizontal length of the well. The subscript sc refers to evaluation at standard condition, 1 bar and 15.6°C. All gas volumes and volumetric rates are given in standard conditions in the rest of the paper. The product  $\mu c$  in (3a) is evaluated at given pressures, hence rendering a semi-linear PDE. This technique is elaborated in Section 3 below. The initial pseudopressure (3d) is calculated by trapezoidal integration of the corresponding values of  $p, \mu$  and  $Z$  which are obtained by gas correlations, see [20].

The IBVP (3) is discretized in space and time in [11], using a finite difference scheme. Spatial discretization applies central difference approximations, i.e. with a second order accuracy. The time discretization uses the backward Euler scheme, or equivalently, a first order orthogonal collocation on finite elements with fixed element lengths. In the completion and design of shale gas wells, a maximum rate  $q^{\text{max}}$  is normally specified based on the surface equipment specifications together with long-term strategic planning of the wells. Moreover, a minimum wellhead pressure is required with respect to the given line pressure. The well rate will initially, or after a shut-in, deliver a peak or plateau rate for some time until the wellhead pressure is equal to the line pressure and the rate starts to decline. Combining these expected features, a simple aggregated well and wellbore model was derived in [11]. Using this aggregated model together with the discretized reservoir proxy model leads to the following simplified single well and reservoir proxy

model.

$$Am_{k+1} = m_k + Bq_{k+1}, \quad (4a)$$

$$m_0 = m_{\text{init}}, \quad (4b)$$

$$q_k = \min \left\{ q^{\text{max}}, \beta \left( m_{k1} - m(e^S p_w) \right) \right\}, \quad (4c)$$

where  $m \in \mathbb{R}^4$  is a vector containing the pseudopressure in each grid block,  $A$  is a tridiagonal matrix,  $B$  is a single-column matrix,  $\beta$  is a constant, possibly skin-dependent well index and  $p_w$  is the wellhead pressure.

### 2.1. Critical gas rate

A critical gas rate  $q_{\text{gc}}$ , sometimes referred to as the minimum rate to lift, can be specified as a lower bound on the flowing gas rate  $q_k$  in order to ensure continuous removal of liquids in the wellbore. The most widely applied model for calculating  $q_{\text{gc}}$  is given by [21], with the onset of liquid loading observed for the majority of gas wells to be controlled by the wellhead conditions. Coleman, Clay, Mccurdy, and Norris III [22] modified Turner's criterion for onset of liquid loading for gas wells with lower wellhead pressures. As shale-gas wells normally operate at low wellhead pressures, we hence use their model for calculating  $q_{\text{gc}}$ . By ensuring that the producing gas rate is always greater or equal to  $q_{\text{gc}}$ , we avoid accumulation of liquids in the vertical part of the wellbore. Furthermore, we limit the valid region of our model and avoid the need to use a multiphase model for low liquid systems as dry and semi-dry shale-gas wells.

## 3. Computing model parameters

The gas-rate profile from shale and fractured tight formation wells is observed to be defined by three distinct sets of transients. The first is a steep decline following the early-time peak or plateau rate, during which the gas initially stored in the fracture network is drained. The second is a very long transient with low rates, often characterized as a pseudo steady-state condition. The last category is pressure build-up transients observable during re-openings of wells after shut-ins. Modeling all of these transients accurately is recognized as inherently difficult [23], even for very complex and detailed numerical models. Consequently, there are fundamental trade-offs when constructing shale-gas models, and clearly a proxy modeling scheme must take into account which type of transients the proxy should replicate.

The basis for the scheme we propose is to perform short, regular shut-ins to prevent liquid loading. Clearly we are most interested in constructing a model replicating the last of the aforementioned categories of transients, i.e. the transients from re-opening a well after a shut-in. During shut-ins, the gas stored in the low permeable shale matrix will flow into the highly conductive fractures and refill these small volumes with gas. The transients in the gas rate when the well is subsequently started up will exhibit slightly different characteristics depending on the length of the shut-ins and the pressures at the time of shut-in. The duration of these transients depends on the pressure build-up in the near-wellbore region, which in turn depends on several factors. The first is the essential reservoir characteristics such as the formation permeability, the conductivity of the fractures and the effective volume of the hydraulically induced and natural fractures. Some of these properties, in particular the formation permeability and the average fracture half-length, can be estimated using the techniques described in [7] and [18]. Other effects

that impact the magnitude of the pressure build-up during shut-ins include the total compressibility in (3a). Assuming negligible rock compressibility and using the equation of state for real gases, the total compressibility for gas reservoirs under isothermal conditions can be expressed as [20]

$$c(p) = \left. \frac{1}{\rho} \frac{\partial \rho}{\partial p} \right|_T = \frac{1}{p} - \frac{1}{Z(p)} \left. \frac{\partial Z(p)}{\partial p} \right|_T. \quad (5)$$

The compressibility of the gas decreases the initial pressure build-up during shut-ins, as some of the pressure change is used to decrease the volume of the gas in pores of the reservoir. These pressure effects are strongly nonlinear, and recognized as difficult to include in semi-analytical and proxy models, particularly in conjunction with the use of pseudopressure  $m(p)$ . Commonly applied techniques include evaluating  $c(p)$  at the initial pressure or at the average of the initial reservoir pressure and the bottomhole pressure. Both of the former techniques is equivalent with assuming a slightly compressible fluid. However, instead of choosing at which pressures to evaluate the compressibility, we parametrize the proxy model (3a) with distinct compressibilities  $c_f$  and  $c_{\text{tran}}$  in the fracture region and the transition region, respectively, and apply a tuning of the model to achieve a good compromise between the compressibility effects during pressure depletion and build-up. As the pressure in the outer low permeability region remains almost constant, we evaluate  $c(p)$  at initial conditions in this region. We further estimate a fracture radius  $r_f$  and an approximated well length  $\tilde{h}$  for describing the varying effective high-conductivity volume in Figure 3, and an average permeability  $k_{\text{tran}}$  in the intermediate transition region. Together, this defines the vector

$$\theta = [r_f \quad k_{\text{tran}} \quad c_f \quad c_{\text{tran}} \quad \tilde{h}]^T \quad (6)$$

of unknown parameters in the proxy model (4).

To estimate parameters in the proxy model, we formulate a constrained weighted nonlinear least squares (WNLS) problem. Using weighted least squares in model tuning and system identification as apposed to ordinary LS estimation, provides a means for emphasizing certain dynamics of the system to be modeled. The weighting of the residuals in the WNLS problem can be obtained by prefiltering of the data [24]. A linear prefilter acts as a frequency weighting of the objective function, and can be designed to increase the model accuracy in the frequency range of interest. This interpretation of the prefilter, however, only applies to strictly linear system identification [25]. Moreover, the equivalence of input-output data filtering and prediction error filtering in linear LS system identification [24], does in general not hold for nonlinear systems [25]. By using the Volterra series representation of nonlinear systems together with the generalized frequency response function, Spinelli et al. [25] demonstrated that reduced estimation bias in the desired bandwidth can be achieved by filtering of the prediction errors compared to separate filtering of the input/output data set. Consequently, even though the frequency weighting interpretation only strictly holds for linear parameter estimation, we can obtain a similar effect in the nonlinear case by filtering the prediction errors through a properly designed filter. Let the prefilter  $L_p(\gamma)$  be given by a general infinite impulse-response filter (IIR),

$$L_p(\gamma) = \frac{C(\gamma)}{D(\gamma)} = \frac{\sum_{m_1=0}^{M_1} c_{m_1} \gamma^{-m_1}}{1 + \sum_{m_2=1}^{M_2} d_{m_2} \gamma^{-m_2}}, \quad (7)$$



where  $\gamma^{-1}$  is the unit delay operator<sup>1</sup>. The filter (7) can be included as a recursive digital filter in the WNLS. Note that the filter equations then simply amounts to a set of linear constraints enforced for each timestep, and hence do not increase the complexity of the resulting NLP. We use a second-order digital low-pass Butterworth filter with a bandwidth  $[0, 0.4]$ , normalized with respect to the Nyquist frequency. Let

$$\epsilon_k := q_k - q_k^{\text{MFR}} \quad (8)$$

be the prediction error, where  $q^{\text{MFR}}$  is the gas rate from the multi-fracture reference model (MFR). Further, let  $\bar{y} = [\bar{y}_1 \bar{y}_2 \dots]^T$  be a predefined vector of *binary* valve-settings defining a shut-in schedule. Using an estimation horizon  $\hat{\mathcal{K}}$ , we then estimate the proxy model parameters  $\theta$  by solving the NLP

$$\min_{\theta} \sum_{k \in \hat{\mathcal{K}}} (\epsilon_k^f)^2 \quad (9a)$$

s.t.

$$\epsilon_k^f = L_p(\gamma)\epsilon_k \quad \forall k \in \{\mathcal{K} : k \geq \max(M_1, M_2)\} \quad (9b)$$

$$\epsilon_k = q_k - q_k^{\text{MFR}} \quad \forall k \in \mathcal{K} \quad (9c)$$

$$A(\theta) m_{k+1} = m_k + B(\theta)q_{k+1}, \quad \forall k \in \mathcal{K} \setminus K \quad (9d)$$

$$m_0 = m_{\text{init}}, \quad (9e)$$

$$q_k = \frac{1}{2} \left( \tilde{q}_k + q^{\text{max}} - \sqrt{(\tilde{q}_k - q^{\text{max}})^2 + \delta_1^2 + \delta_2^2} \right), \quad \forall k \in \mathcal{K} \quad (9f)$$

$$\tilde{q}_k = \bar{y}_k \beta(\theta) \left( m_{k1} - m(e^S p_w) \right), \quad \forall k \in \mathcal{K} \quad (9g)$$

$$\theta^{\text{lo}} \leq \theta \leq \theta^{\text{up}} \quad (9h)$$

$$c_{\text{tran}} \leq c_f \quad (9i)$$

where (9f) is a smooth approximation of the non-smooth well model (4c), with smoothing parameters  $\delta_1$  and  $\delta_2$  selected from the interval  $[10^{-2}, 10^{-3}]$ . The second smoothing parameter  $\delta_2$  is added to prevent negative rates when the well is shut-in. Observe that whenever  $\bar{y}_k = 0$  for some  $k \in \mathcal{K}$ , then  $\tilde{q}_k = 0$  and hence  $q_k \approx 0$  since (9f) gives the minimum of  $\tilde{q}$  and  $q^{\text{max}}$ . The constraint (9i) is based on the property that average compressibility in the reservoir is higher closer to the wellbore, and serves as a regularization of the parameter estimation problem in addition to the bounds (9h) [26; 27]. Note that the compact parametric representation (9d) of the discretized PDE model consists of a set of nonlinear equations used to define the spatial gridding and the transmissibilities in the reservoir model. For further details on the derivation of the proxy model (9d), see [11].

For the tuning of the proxy model, we use a high-fidelity multi-fracture reference model constructed with cylindrical geometry similar to Figure 3 with ten equally spaced fractures. The high-fidelity model is implemented in the state-of-the-art simulation software SENSOR [28], using logarithmic grid refinements between and outside the fractures, with totally 4200 grid cells. We further extend the high-fidelity reservoir model with a static wellbore model, see e.g.

<sup>1</sup>The unit delay operator is usually defined as  $q^{-1}$  [24]. However, as  $q$  is used as notation for the gas rate, we use for notational convenience  $\gamma^{-1}$  as the equivalent operator.

[29]. Essential given model parameters in the proxy and the reference model are quoted in Table 1. We use a  $\Delta k = 1$  day fixed timestep for the proxy model, while we use a one hour timestep for the multi-fracture reference model to capture more of the high-frequency dynamics in this model. Note that in practice the rates are likely to be available on a daily basis only. The NLP (9) is implemented in GAMS [30] and solved with the NLP solver CONOPT v3.15. The results of the tuning of the proxy model is shown in Figure 5, where we have used two shut-ins of five and three days, respectively, to excite the model. The proposed modeling and tuning scheme renders a good match of transients after shut-ins of a couple of days. The match is observed to be tighter in the pseudo-steady state region for the second transient compared to transient after the first (five-day) shut-in, while a discrepancy is seen in the peak-rate after the second, three-day long shut-in. This discrepancy reflects the trade-offs made in the proxy modeling (4) and in the design of the prefilter (7): increasing the bandwidth or shifting the bandpass would increase the gain on the fast modes of the system and hence, give a better match of the second peak in Figure 5. A nonzero steady-state gain is necessary however to obtain an acceptable match of the models at the time at which the rates crossed  $q_{gc}$ , as the rate at 585 days are almost steady-state. Arguably, one could consider using a bandpass filter instead of a low-pass filter to place more emphasis on the transient modes and thereby better match the peak rates. This will, however, give a higher bias in the rate of the proxy model in the pseudo steady-state regions, and lead to an almost non-decreasing rate after the initial decline of the rate. Some of the mismatch in the curvature of the transients is also due to the constant compressibility assumption, while parts of the mismatch may be compensated for by a higher-order discretization scheme. Note that the parameter estimate will be somewhat sensitive to the NLP starting point since the WLNS (9) is a non-convex NLP. To compensate for this dependency, we have used a multi start-point strategy with different initial choices  $\theta^{\text{init}}$  to provide good starting points for the NLP (9).

The advantage of using a Butterworth filter is the flat magnitude in the bandpass. However, as Butterworth filters are all-pole filters, there will be a significant phase shift at the output of the filter, particularly for high-order filters. Care must therefore be taken with respect to the choice of filter order if the length of the time-series of production rates is short and there are shut-ins at the end of time-series, as parts of the transient may be lost due to the phase lag of the filter. Compared to linear system identification, in which a prefilter tends to cut off model content in frequencies outside the bandpass, the prefilter applied in nonlinear system identification introduces a scattering of the frequency spectrum outside the bandpass. This property may impact the frequency weighting in (9) and hence the model match. See [25] for further details and analysis.

Table 1: Given reservoir model parameters.

Symbol	Parameter	Value	Unit
$\phi$	porosity	6	%
$k_m$	formation permeability	$3 \times 10^{-4}$	mD
$z_w$	true vertical depth	2300	m
$p_r$	initial reservoir pressure	200	bar
$p_w$	minimum wellhead pressure	6.9	bar
$q^{\text{max}}$	maximum rate	$4 \times 10^4$	m <sup>3</sup> /d
$h$	true length of horizontal wellbore	492	m
$n_f$	true number of hydraulic fractures	10	-

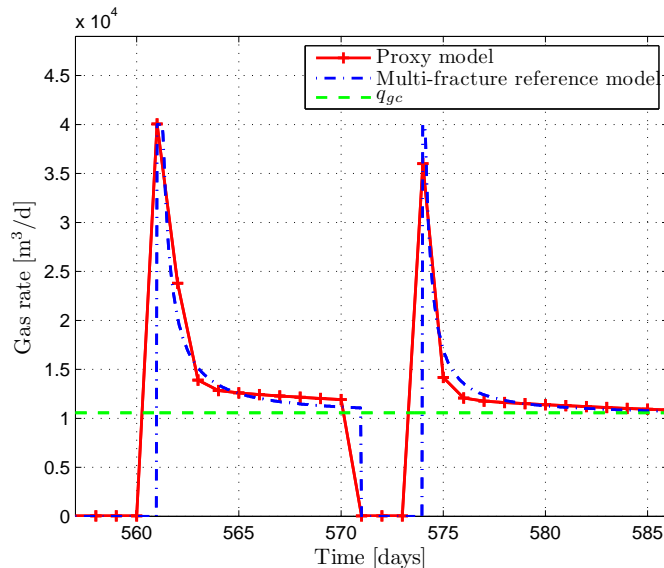


Figure 5: Tuning of parameters  $\theta$  the proxy model. The first transient corresponds to a five-day shut-in, while the second shut-in is three days long.

### 3.1. Updating model parameters

One of the key issues for industrial appreciation of model-based decision support tools is the user's ability to maintain and update the models used [31]. This is particularly challenging for asset and field-wide optimization similar to the problem in this paper, which includes multiple disaggregated model types such as reservoir, surface and wells models, all having different time scales. Challenges are related to the complexity of the models used in the optimization [32], the level of automation of the model update, and the (possibly excessive) amount of data available. Quite commonly, models are updated quite infrequently, e.g. on a yearly basis [31]. Such seldom model updates introduce a significant time-delay in the feedback decision loop, and clearly deteriorate the accuracy and the performance of the decision support tool [31]. In this context, there is an evident advantage of using model predictive control (MPC) compared to real-time optimization with separate model updating, as model errors and slowly drifting in parameters are compensated for by the per-timestep feedback and estimation of initial conditions. However, the use of MPC in a large, geographically spread petroleum production system introduces several challenges, and requires a high level of automatic control and less human-in-the-loop operations.

Although a cross-validation in [10] of a similar realization of the proxy model (4) showed a good match with the reference model, the accuracy of the proxy model will abate over time. This is a results of the tuning technique (9d) in which only a limited window of data and shut-ins is used, as well as the prefilter. On a long horizon, though, the actual flow and pressure characteristics of the reservoir change as result of local pressure depletion in the fracture network, transition to boundary dominated flow and change in proppant distribution and near-wellbore skin [18]. Shale-gas reservoirs may also contain time-varying parameters such as pressure dependent formation permeability [16]. The parameter vector  $\theta$  should hence be updated using a moving horizon-type (MHE) parameter estimation technique [33]. Using a window length  $\mathcal{K}_u$  of only the last set of applied shut-ins, we limit the weight on the earlier observed shut-ins and

tuning of the proxy model. For a given  $\theta^*$  obtained by solving the initial parameter estimation (9) shown in Fig. 5, we update the proxy model by solving an augmented form of the WNLS problem (9),

$$\min_{\theta} \sum_{k \in \mathcal{K}_u} (\epsilon_k^f)^2 + (\theta - \theta^*)^T D_s^T P D_s (\theta - \theta^*) \quad (10a)$$

s.t.

$$\text{equation (9b) - (9i)}, \quad (10b)$$

where  $P$  is a symmetric positive definite matrix, and  $D_s$  is a diagonal scaling matrix used to normalize the different parameters in  $\theta$ . The updating of the proxy model by solving (10) is qualitatively different from ordinary moving horizon based parameter (and state) estimation [33]. In MHE, the quadratic term in (10a) corresponds to an arrival cost on new data, where an optimal  $P$  can be computed by an Extended Kalman filter or other techniques. Consequently,  $P$  is in this case used to aggregate previously obtained information on the model. The shale-well proxy model, however, is only valid in a limit area (i.e. over a certain time and pressure range). Rather than aggregating information, we want the estimates to be gradually forgotten as new shut-in data becomes available. Consequently, we update the matrix  $P$  as a type of forgetting factor [24],

$$P^{\hat{k}+1} = \alpha I P^{\hat{k}} \quad (11)$$

and use the last estimate  $\hat{\theta}$  as the starting point when solving the NLP (10). Figure 6 compares

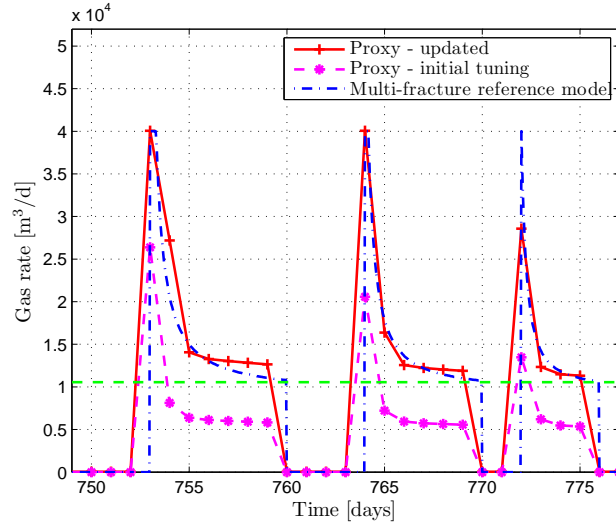


Figure 6: Updating parameters in the proxy model.

the rate of the proxy model using the initial optimal parameter vector  $\theta^*$  with the updated proxy model from solving (10). The initially tuned proxy preserves the essential curvature in the rate transients, but displays a prominent bias. This bias relates to the total pore volume of the proxy model compared to the actual reservoir (reference model). The volumes of the fracture and transition region are too small to capture the long-term gas drainage as they compensate for

short-term pressure transients during shut-ins, hence leading to a fast pressure depletion. The outer, low permeability grid block, which in a way acts as a source term, is unable to compensate for this pressure depletion, hence leading to high pressure gradients. Figure 6 also shows the difficulty encountered when shorter shut-ins are used to update the proxy as seen after 772 days. For such short shut-ins the dynamics are even faster, and the proxy does not match the high peak after re-opening the well.

#### 4. Formulation of optimization problem

Table 2: Indices and set definitions

Index	Interpretation	Set	Elements
$i$	spatial reservoir grid block	$\mathcal{I}$	$\{1 \dots I\}$
$j$	well number	$\mathcal{J}$	$\{1 \dots J\}$
$k$	discrete time index	$\mathcal{K}$	$\{0 \dots K\}$
$l$	pad number	$\mathcal{L}$	$\{1 \dots L\}$
$n$	iteration in Lagrangian scheme	-	$\{1 \dots N\}$
$d$	terms in the disjunctions	-	$\{1, 2, 11, 12, 13\}$

This section describes the formulation of the optimization problem for the shale-gas multi-pad structure depicted in Figure 1. We assume that a production plan for  $K$  days in terms of a reference rate  $q^{\text{REF}}$  is given from strategic and contract based management of the field. Without loss of generality, we will assume that the reference rate is constant for the entire horizon  $\mathcal{K}$ . When shutting in multiple wells at a single location without some scheduling, then in the worst case the entire set of wells at one pad may be shut in at the same time giving zero total rate, while the peak rate when the wells are subsequently re-opened may be unfeasibly high. This must clearly be avoided. We want each pad  $l \in \mathcal{L}$  to produce a stabilized rate  $q_l^{\text{ref}}$  to prevent high pressure and rate oscillations in the feed to the compressors, while at the same time, produce close to the overall reference rate  $q^{\text{REF}}$  over the given planning horizon.

Optimization problems with logical structures and conditions may benefit from starting with a generalized disjunctive programming (GDP) formulation [34] instead of an ad-hoc mixed integer formulation, as a GDP formulation generally captures more directly the connections between the logical part and the constraints of the problem [35]. Based on a nonconvex and nonsmooth GDP formulation using the well model (4c) explicitly in the disjunction, an MILP reformulation of the problem was derived in [10]. The reformulation is based on a direct big-M type reformulation of the disjunction with an subsequent exact reformulation of the min-function, and was shown to be computationally superior to equivalent reformulations of the disjunction with smooth approximations of the min-function. The min-function, however, is a disjunction itself [36]. Consequently, we can rewrite the nonsmooth well-model (4c) as a disjunction based on the well inflow model and the maximum rate  $q^{\text{max}}$ . This disjunction is then embedded into the main disjunction deciding if a well is producing or shut in. We further include logical constraints for requiring a minimum shut-in and production time between each succeeding shut-in cycle [10]. By combing the reference rate objectives described in the previous paragraph with the logical

formulation for shutting in the wells, we obtain the GDP model

$$Z = \min \sum_{l \in \mathcal{L}} \max_k |q_l^{\text{ref}} - q_{lk}^{\text{pad}}| \quad (12a)$$

s.t.

$$\sum_{l \in \mathcal{L}} q_l^{\text{ref}} = q^{\text{REF}} \quad (12b)$$

$$q_{lk}^{\text{pad}} = \sum_{j \in \mathcal{J}} q_{jk}, \quad \forall l \in \mathcal{L}, \forall k \in \mathcal{K} \quad (12c)$$

$$A_{lj} m_{ljk+1} = m_{ljk} + B_{lj} q_{ljk+1}, \quad \forall l \in \mathcal{L}, \forall j \in \mathcal{J}, k \in \mathcal{K} \setminus K \quad (12d)$$

$$m_{lj0} = m_{lj}^{\text{init}}, \quad \forall l \in \mathcal{L}, \forall j \in \mathcal{J} \quad (12e)$$

$$\tilde{q}_{ljk} = \beta_j (m_{ljk1} - m(e^{S_{lj}} p_{w,lj})), \quad \forall l \in \mathcal{L}, \forall j \in \mathcal{J}, \forall k \in \mathcal{K} \quad (12f)$$

$$\left[ \begin{array}{c} Y_{ljk}^1 \\ q_{ljk} \geq q_{\text{gc}} \\ Y_{ljk}^{11} \\ \tilde{q}_{ljk} \leq q_{\text{max}} - \delta^e \\ q_{ljk} = \tilde{q}_{ljk} \end{array} \right] \vee \left[ \begin{array}{c} Y_{ljk}^{12} \\ \tilde{q}_{ljk} \geq q_{\text{max}} \\ q_{ljk} = q_{\text{max}} \end{array} \right] \vee \left[ \begin{array}{c} Y_{ljk}^2 \\ q_{ljk} \equiv 0 \end{array} \right], \quad \forall l \in \mathcal{L}, \forall j \in \mathcal{J}, k \in \mathcal{K} \quad (12g)$$

$$Y_{ljk}^1 \vee Y_{ljk}^2, \quad \forall l \in \mathcal{L}, \forall j \in \mathcal{J}, k \in \mathcal{K} \quad (12h)$$

$$Y_{ljk}^1 \Leftrightarrow Y_{ljk}^{11} \vee Y_{ljk}^{12}, \quad \forall l \in \mathcal{L}, \forall j \in \mathcal{J}, k \in \mathcal{K} \quad (12i)$$

$$(Y_{ljk-1}^1 \wedge \neg Y_{ljk}^2) \Rightarrow Y_{ljk+1}^2 \wedge \dots \wedge Y_{ljk+\tau_1-1}^2, \quad \forall l \in \mathcal{L}, \forall j \in \mathcal{J}, k \in \{\mathcal{K} : 1 \leq k \leq K - \tau_1 + 1\} \quad (12j)$$

$$(Y_{ljk}^1 \wedge Y_{ljk-1}^2) \Rightarrow Y_{ljk+1}^1 \wedge \dots \wedge Y_{ljk+\tau_2-1}^1, \quad \forall l \in \mathcal{L}, \forall j \in \mathcal{J}, k \in \{\mathcal{K} : 1 \leq k \leq K - \tau_2 + 1\} \quad (12k)$$

$$Y_{ljk}^1, Y_{ljk}^{11}, Y_{ljk}^{12}, Y_{ljk}^2 \in \{\text{True}, \text{False}\}$$

Observe that in (12a) we have used the sum of the infinity norm between the per timestep pad rate  $q_{lk}^{\text{pad}}$  and the corresponding pad reference rate  $q_l^{\text{ref}}$  as the objective function. The constant  $\delta^e$  in (12g) is included to enforce non-satisfaction of both of the terms in the embedded disjunctions, and is set typically to 0.001 in implementations (cf. [37]). The above logic-based problem formulation is in non-standard GDP form [34; 35] due to the embedded disjunction (12g), but can be systematically transformed into standard GDP form by using the technique described

[35; 37]. This leads to the following logically equivalent representation of the disjunction (12g),

$$\left[ \begin{array}{c} Y_{ljk}^1 \\ q_{ljk} \geq q_{gc} \end{array} \right] \vee \left[ \begin{array}{c} Y_{ljk}^2 \\ q_{ljk} = 0 \end{array} \right], \quad \forall l \in \mathcal{L}, \forall j \in \mathcal{J}, k \in \mathcal{K} \quad (13a)$$

$$\left[ \begin{array}{c} Y_{ljk}^{11} \\ \tilde{q}_{ljk} \leq q^{\max} - \delta^e \\ q_{ljk} = \tilde{q}_{ljk} \end{array} \right] \vee \left[ \begin{array}{c} Y_{ljk}^{12} \\ \tilde{q}_{ljk} \geq q^{\max} \\ q_{ljk} = q^{\max} \end{array} \right] \vee \left[ Y_{ljk}^{13} \right], \quad \forall l \in \mathcal{L}, \forall j \in \mathcal{J}, k \in \mathcal{K} \quad (13b)$$

$$Y_{ljk}^1 \vee Y_{ljk}^{11}, \quad \forall l \in \mathcal{L}, \forall j \in \mathcal{J}, k \in \mathcal{K} \quad (13c)$$

$$Y_{ljk}^{11} \vee Y_{ljk}^{12} \vee Y_{ljk}^{13}, \quad \forall l \in \mathcal{L}, \forall j \in \mathcal{J}, k \in \mathcal{K} \quad (13d)$$

$$Y_{ljk}^1 \Leftrightarrow Y_{ljk}^{11} \vee Y_{ljk}^{12}, \quad \forall l \in \mathcal{L}, \forall j \in \mathcal{J}, k \in \mathcal{K} \quad (13e)$$

$$Y_{ljk}^{13} \Leftrightarrow Y_{ljk}^{12}. \quad \forall l \in \mathcal{L}, \forall j \in \mathcal{J}, k \in \mathcal{K} \quad (13f)$$

The Boolean  $Y^{13}$  is introduced to allow the logical state  $\neg Y^{11} \wedge \neg Y^{12}$ . Transforming (12) into the standard form (13) allows us to assess different ways of reformulating the disjunctions into algebraic equations. Raman and Grossmann [34] showed how linear GDPs can be converted to MILPs by introducing binary variables  $y$  with one-to-one correspondence with the Boolean variables  $Y$ , and using either a Big-M reformulation [36] or the convex hull description of general linear disjunctions derived by Balas [38]. The polyhedral convex hull description of linear disjunctions is based on disaggregation of the variables in the disjunction and assuring that only one of the terms in the disjunction is active. The convex hull reformulation yields an LP relaxation at least as tight or tighter than big-M reformulations, and hence generally stronger lower bounds and reduced computation time [35; 39]. Still, the convex hull reformulation requires more variables and constraints than big-M reformulations. This may translate into longer computation time for some problems [35; 39]. Based on our computational experience, we observed that a full convex hull reformulation of the disjunctions (13a)–(13b) resulted in longer computation times than using the convex hull reformulation for (13a) and a big-M reformulation of (13b), while only a marginal improvement in the LP relaxation was observed when using the full convex hull reformulation. Disaggregation of  $q_{ljk}$  in (13a) then yields

$$q_{ljk} = q_{ljk}^1 + q_{ljk}^2, \quad \forall l \in \mathcal{L}, \forall j \in \mathcal{J}, k \in \mathcal{K} \quad (14)$$

$$y_{ljk}^1 + y_{ljk}^2 = 1, \quad \forall l \in \mathcal{L}, \forall j \in \mathcal{J}, k \in \mathcal{K} \quad (15)$$

$$q_{ljk}^1, q_{ljk}^2 \geq 0, \quad (16)$$

which by substitution of the disaggregated variables into (13a) simplifies to

$$q_{gc} y_{ljk}^1 \leq q_{ljk} \leq y_{ljk}^1 q^{\max}, \quad \forall l \in \mathcal{L}, \forall j \in \mathcal{J}, k \in \mathcal{K} \quad (17)$$

since  $q_{ljk}^2 \equiv 0$ , and  $\tilde{q}_{ljk}$  does not appear in this disjunction. The big-M reformulation of (13b) is obtained by rewriting the equality constraints as two inequalities and introducing big-M parameters for each term. The logical proposition (12j)–(12k) for the minimum shut-in and production time can be transformed to linear algebraic inequalities by replacing the implications with disjunctions, converting the resulting disjunction to conjunctive normal form and then replacing the logical relations with its algebraic counterparts, see [40]. This eventually leads to the following primal MILP problem.

$$Z = \min \sum_{l \in \mathcal{L}} f_l \quad (18a)$$

s.t.

$$q^{\text{REF}} = \sum_{l \in \mathcal{L}} q_l^{\text{ref}} \quad (18b)$$

$$f_l \geq q_l^{\text{ref}} - q_{lk}^{\text{pad}}, \quad \forall l \in \mathcal{L}, k \in \mathcal{K} \quad (18c)$$

$$f_l \geq -q_l^{\text{ref}} + q_{lk}^{\text{pad}}, \quad \forall l \in \mathcal{L}, k \in \mathcal{K} \quad (18d)$$

$$q_{lk}^{\text{pad}} = \sum_{j \in \mathcal{J}} q_{jk}, \quad \forall l \in \mathcal{L}, \forall k \in \mathcal{K} \quad (18e)$$

$$A_{lj} m_{ljk+1} = m_{ljk} + B_{lj} q_{ljk+1}, \quad \forall l \in \mathcal{L}, \forall j \in \mathcal{J}, k \in \mathcal{K} \setminus K \quad (18f)$$

$$m_{lj0} = m_{\text{init}}, \quad \forall l \in \mathcal{L}, \forall j \in \mathcal{J} \quad (18g)$$

$$\tilde{q}_{ljk} = \beta_j (m_{ljk1} - m(e^{S_{ij}} p_{w,lj})), \quad \forall l \in \mathcal{L}, \forall j \in \mathcal{J}, \forall k \in \mathcal{K} \quad (18h)$$

$$q_{\text{gc},ljk}^1 \leq q_{ljk}, \quad \forall l \in \mathcal{L}, \forall j \in \mathcal{J}, k \in \mathcal{K} \quad (18i)$$

$$q_{ljk} \leq q^{\text{max}} y_{ljk}^1, \quad \forall l \in \mathcal{L}, \forall j \in \mathcal{J}, k \in \mathcal{K} \quad (18j)$$

$$\tilde{q}_{ljk} - q^{\text{max}} + \delta^e \leq M^{11} (1 - y_{ljk}^{11}), \quad \forall l \in \mathcal{L}, \forall j \in \mathcal{J}, k \in \mathcal{K} \quad (18k)$$

$$q_{ljk} - \tilde{q}_{ljk} \leq M^{11} (1 - y_{ljk}^{11}), \quad \forall l \in \mathcal{L}, \forall j \in \mathcal{J}, k \in \mathcal{K} \quad (18l)$$

$$\tilde{q}_{ljk} - q_{ljk} \leq M^{11} (1 - y_{ljk}^{11}), \quad \forall l \in \mathcal{L}, \forall j \in \mathcal{J}, k \in \mathcal{K} \quad (18m)$$

$$q^{\text{max}} - \tilde{q}_{ljk} \leq M^{12} (1 - y_{ljk}^{12}), \quad \forall l \in \mathcal{L}, \forall j \in \mathcal{J}, k \in \mathcal{K} \quad (18n)$$

$$q_{ljk} - q^{\text{max}} \leq M^{12} (1 - y_{ljk}^{12}), \quad \forall l \in \mathcal{L}, \forall j \in \mathcal{J}, k \in \mathcal{K} \quad (18o)$$

$$q^{\text{max}} - q_{ljk} \leq M^{12} (1 - y_{ljk}^{12}), \quad \forall l \in \mathcal{L}, \forall j \in \mathcal{J}, k \in \mathcal{K} \quad (18p)$$

$$y_{ljk}^{11} + y_{ljk}^{12} + y_{ljk}^{13} = 1 \quad \forall l \in \mathcal{L}, \forall j \in \mathcal{J}, k \in \mathcal{K} \quad (18q)$$

$$y_{ljk}^{11} + y_{ljk}^{12} = y_{ljk}^1 \quad \forall l \in \mathcal{L}, \forall j \in \mathcal{J}, k \in \mathcal{K} \quad (18r)$$

$$y_{ljk}^1 + y_{ljk}^2 = 1 \quad \forall l \in \mathcal{L}, \forall j \in \mathcal{J}, k \in \mathcal{K} \quad (18s)$$

$$y_{ljk}^{13} = y_{ljk}^2 \quad \forall l \in \mathcal{L}, \forall j \in \mathcal{J}, k \in \mathcal{K} \quad (18t)$$

$$y_{ljk-1}^1 + y_{ljk}^2 \leq 1 + y_{lj\rho}^2, \quad \forall l \in \mathcal{L}, \forall j \in \mathcal{J}, k \in \mathcal{K} \setminus 0 \quad (18u)$$

$$\rho \in [k + 1, \min\{k + \tau_1 - 1, K\}] \quad (18v)$$

$$y_{ljk-1}^2 + y_{ljk-1}^1 \leq 1 + y_{lj\rho}^1, \quad \forall l \in \mathcal{L}, \forall j \in \mathcal{J}, k \in \mathcal{K} \setminus 0 \quad (18w)$$

$$\rho \in [k + 1, \min\{k + \tau_2 - 1, K\}] \quad (18x)$$

Introducing the auxiliary variables  $f_l$  with the constraints (18c)–(18d) exactly reformulates the non-smooth objective function (12a). Note that although the MILP formulation (18) can be considerably reduced by substituting dependent binary variables, we leave this reduction and tightening of the problem to the presolve routines in the MILP solver (such as CPLEX and Gurobi).

Similar constraints as (18u)–(18w) for the minimum shut-in and production time for each



shut-in cycle have been used for defining minimum up and down time requirements in the hydro-thermal unit commitment (UC) problem in power generation [41], as well as restricted length set-up sequences in single-item lot sizing [42]. Several authors have studied the tightness of different formulations of these constraints in the context of the UC problem [43; 44]. Rajan and Takriti [44] derives the convex hull polytope of an extended variable formulation of the minimum up and downtime constraints, in which start-up and shut-down costs are present. As the problem (12) contains no costs on the shut-ins and start-ups, however, we can omit additional binary variables for modeling start-up and shut-ins transitions, and limit the formulation of the minimum shut-in and production time constraints using only the on/off variables  $y_{ljk}^1$  and  $y_{ljk}^2$  with the constraints (18u)–(18w). Note that these constraints are projections of the so-called turn on/off inequalities derived in [44] onto the space of the  $y_{ljk}^1$  and  $y_{ljk}^2$  variables [42].

## 5. Lagrangian relaxation scheme

Lagrangian relaxation is an efficient solution technique for problems with a block-separable structure in the constraint set and few binding or coupling constraints. This structure is often found in optimization problems with some sort of network flow and spatial distribution [12; 13]. The primal problem (18) clearly renders this type of problem structure, as the only constraint that links the  $\mathcal{L}$  pads together is the constraint (18b) requiring that the sum of the individual pad reference rates must equal the total requested rate  $q^{\text{REF}}$ . Consequently, by dualizing this constraint, we will obtain a block-separable, decomposable problem with one subproblem for each pad  $l \in \mathcal{L}$ . Let  $\lambda \in \mathbb{R}$  be the Lagrangian multiplier associated with the reference rate equality constraint (18b). The Lagrangian relaxation is then given by

$$Z_{\text{LR}}(\lambda) = \min \sum_{l \in \mathcal{L}} f_l + \lambda \left( \sum_{l \in \mathcal{L}} q_l^{\text{ref}} - q^{\text{REF}} \right) \quad (19)$$

s.t.  
eq. (18c)–(18w),

hence rendering a spatial decomposition over the set  $\mathcal{L}$ , such that solving the Lagrangian (19) means solving  $l$  independent subproblems defined by

$$Z_{\text{LR}}^l = \min f_l + \lambda q_l^{\text{ref}} \quad (20)$$

s.t.  
eq. (18c)–(18w) for given  $l \in \mathcal{L}$ .

For any  $\lambda \in \mathbb{R}$ , the Lagrangian (19) is a relaxation of the primal problem (18), and consequently defines a lower bound on  $Z$  [45; 46]. The solution of the Lagrangian (19) yields a lower bound on  $Z$  as least as tight as the LP relaxation [45]. The dual variable  $\lambda$  is a marginal cost on the demand constraint (18b) of the pads' reference rate  $q_l^{\text{ref}}$  relative to the given total reference rate (or demand)  $q^{\text{REF}}$ , by pricing the cost of satisfying an additional unit in total gas demand [13].

Obtaining the best possible value  $Z_{\text{LR}}(\lambda)$  of the Lagrangian relaxation (19), that is, the one providing the tightest lower bound on  $Z$ , requires finding the optimal multiplier  $\lambda$ . This problem defines the Lagrangian Dual [45],

$$Z_{\text{D}} = \max_{\lambda} Z_{\text{LR}}(\lambda). \quad (21)$$

The Lagrangian  $Z_{\text{LR}}(\lambda)$  can be shown to be concave and piecewise linear, and hence nonsmooth [36]. Solving (21) hence requires using techniques for nonsmooth optimization. The solution of (21) will in most cases be primal infeasible in the sense that the dualized reference rate constraint (18b) will be violated. Consequently, some technique using the information from solving the Lagrangian relaxation and/or the Lagrangian dual must be developed for generating primal feasible solutions.

### 5.1. Solving the Lagrangian Dual

The two most common classes of algorithms for solving the Lagrangian duals are subgradients methods and methods based on the cutting-plane approach [47]. A variety of implementations and modification of algorithms within these two classes exists. The subgradient-type methods do not require solution of any optimization problem and are hence easy to implement. They are based on using the dualized constraint, i.e. (18b) in the Lagrangian relaxation (19), as the subgradient for  $Z_{\text{LR}}$  in the space of the dual variables. Subgradient methods may require extensive tuning of the stepsize parameters to obtain good practical convergence, and lack a true termination criteria. By substituting (19) for  $Z_{\text{LR}}$  in the Lagrangian Dual (21), we obtain a maximin problem, which can be equivalently written as the semi-infinite linear program [48]

$$Z_{\text{D}} = \max_{\lambda, \eta} \eta \quad (22a)$$

s.t.

$$\eta \leq \sum_{l \in \mathcal{L}} f_l + \lambda \left( \sum_{l \in \mathcal{L}} q_l^{\text{ref}} - q^{\text{REF}} \right), \quad (f_l, q_l^{\text{ref}}) \in Q \quad (22b)$$

where  $Q$  is a polytope, by assumption, defining the feasible set of the Lagrangian relaxation, and  $\eta \in \mathbb{R}$ . The cutting plane method is based on replacing (22b) with a subset of cuts accumulated by solving the Lagrangian relaxation. This approach renders the linear program

$$Z_{\text{D}} = \max_{\lambda, \eta} \eta \quad (23a)$$

s.t.

$$\eta \leq \sum_{l \in \mathcal{L}} f_l^n + \lambda \left( \sum_{l \in \mathcal{L}} q_l^{\text{ref},n} - q^{\text{REF}} \right) \quad n = 1, 2, \dots \quad (23b)$$

for the Lagrangian Dual, where  $n$  is the number of iterations in the primal-dual loop solving (19) and (23). At each step of the loop, solving (23) provides an upper and lower bound on  $Z_{\text{D}}$  [46],

$$\eta^n \geq Z_{\text{D}} \geq Z_{\text{LR}}^n \quad (24)$$

The method terminates when these bounds coincide. The cutting-plane method has finite convergence [48], but suffers in its basic form from an inherent instability. In early iterations, the cutting-plane model (23b) of  $Q$  is insufficient to provide a bounded problem. Moreover, the sequence  $\lambda^n$  of dual solutions lack local properties [46]. These characteristics often leads to oscillatory performance and slow convergence of the cutting-plane method in practice. The problem of early oscillations and unbounded solutions of (23) can be addressed by adding bounds on the dual variables [48], and adding known solutions  $(\hat{f}_l, \hat{q}_l^{\text{ref}})$  of (18) and (20) to improve the cutting-plane model (23b) of  $Q$  [46]. Many rigorous stabilization methods have been suggested

to increase the local properties and mitigate the oscillations of the dual solutions from the cutting-plane method, see [48]-Ch. XV and [46] for a thorough review. The above approaches for stabilizing the cutting plane method for the Lagrangian Dual may be combined and tailored to derive efficient algorithms for updating the Lagrangian multipliers, or possibly combined with a subgradient method [49].

To stabilize the cutting-plane formulation for solving the Lagrangian Dual (23), we adopt a trust-region scheme from [48] and [50]. The trust-region approach provides a simple device for coping with instabilities in the cutting-plane method when applied to Lagrangian Duals with relatively few dual variables, and is closely related to the boxstep method [51]. The algorithm ensures that the next  $\lambda^{n+1}$  is never further away than a distance  $\Delta$  from the current *stability center*  $\bar{\lambda}$ . The algorithm is as follows:

**Algorithm 1:** Trust-region cutting-plane method

*Initialization:* Select initial trust-region  $\Delta^1$ , stability center  $\bar{\lambda}$ , a termination tolerance  $\rho^{\text{CP}}$ , and a descent coefficient  $\sigma \in (0, 1)$ . Set  $\lambda^1 = \bar{\lambda}$ ,  $\tilde{N} := 1$ , and solve (19) to obtain  $Z_{\text{LR}}(\lambda^1)$ .

*Step 1:* Solve the LP

$$\max_{\lambda, \eta} \eta \quad (25a)$$

s.t.

$$\eta \leq \sum_{l \in \mathcal{L}} f_l^n + \lambda \left( \sum_{l \in \mathcal{L}} q_l^{\text{ref}, n} - q^{\text{REF}} \right), \quad \forall n \leq \tilde{N} \quad (25b)$$

$$|\lambda - \bar{\lambda}| \leq \Delta^n \quad (25c)$$

$$\lambda \geq -1 \quad (25d)$$

to obtain  $\lambda^{n+1}$ . Compute

$$\tilde{\rho} := \eta - Z_{\text{LR}}(\bar{\lambda}) \quad (26)$$

*Step 2:* If  $\tilde{\rho} < \rho^{\text{CP}}$ , terminate with  $\lambda = \bar{\lambda}$  and  $Z_{\text{D}} = Z_{\text{LR}}(\bar{\lambda})$

*Step 3:* Compute  $Z_{\text{LR}}(\lambda^{n+1})$ .

*Step 4:* If

$$Z_{\text{LR}}(\lambda^{n+1}) \geq Z_{\text{LR}}(\bar{\lambda}) + \sigma \tilde{\rho}, \quad (27)$$

update the center:  $\bar{\lambda}^{n+1} = \lambda^{n+1}$ . Otherwise, leave center unchanged.

*Step 5:* Update trust-region; compute the ratio

$$\rho := \frac{Z_{\text{LR}}(\lambda^{n+1}) - Z_{\text{LR}}(\bar{\lambda})}{\tilde{\rho}} \quad (28)$$

If  $\rho = 1$ , set  $\Delta^{n+1} = 1.5\Delta^n$ . If  $\rho < 0$ , set  $\Delta^{n+1} = 0.8\Delta^n$

Set  $n = n + 1$ ,  $\tilde{N} = n + 1$ , and repeat from *Step 1*.

By evaluating in step 4 the ratio of the *predicted* decrease in the gap  $\tilde{\rho}$  with the actual decrease, an ascent-step (serious step) is only declared if a sufficient decrease is obtained relative to the parameter  $\sigma$ . Otherwise, the current iteration is a null-step, while the cutting-plane model (25) is updated. Comparing with the subgradient method, the above scheme guarantees an increase for each serious step. The numerical value of the Armijo-like parameter  $\sigma$  is set to 0.01. The trust-region radius is only updated when  $\lambda$  is on the boundary of the region (cf. [50]). This way of updating the radius, together with the trust-region adjustment factors 0.8, 1.5, may be tuned to improve the performance of the algorithm. Note that we have used the 1-norm for the trust-region (25c); for higher dimensional problems, the Euclidean norm may be more suitable, however leading to a nonlinear problem.

The lower bound (25d) is added to ensure values of  $\lambda$  that gives bounded Lagrangian subproblems (20). The proof of the numerical value of this bound is given in Appendix A. By a similar proof, it can be shown by using Fourier-Motzkin elimination that for a *single* reference rate  $q^{\text{REF}}$ , any non-negative value of  $\lambda$  will cause  $q_l^{\text{ref}} = 0$ , and consequently the optimal solution  $f_l = q_l^{\text{pad}} = 0$  for each of the subproblems (20). Hence, we can add this a priori solution to (25) to improve the initial cutting-plane model of (22). Note that in practice, though, it may be sufficient to select an initially small trust-region  $\Delta$  together with a good estimate for  $\lambda^1 = \bar{\lambda}$  to prevent optimal values of  $\lambda$  less than -1.

## 5.2. Generating primal feasible solutions

Obtaining primal feasible solutions in a Lagrangian relaxation scheme normally requires the use of some, possibly problem specific, heuristic. Either or both of the solutions from the Lagrangian relaxation and the Lagrangian Dual may be exploited to generate primal feasible solutions, often using some type of greedy algorithm. The solutions from the Lagrangian relaxation often becomes almost primal feasible (with respect to the dualized constraints) with increasing number of iterations. As  $Z_{\text{LR}}$  is a best-bound on  $Z$ , the solution of the Lagrangian relaxation may in some way aid the search for primal feasible solutions.

Solving the  $|\mathcal{L}|$  Lagrangian subproblems generates a schedule (or sequence) of shut-ins for each well on the given pad  $l$  together with a reference-rate  $q_l^{\text{ref}}$ . The shut-in schedule is completely described by the value of the binary variables  $y_{ljk}^d$ . Consequently, the solution of the subproblems avoid liquid loading in the wells. On the other hand, the shut-in schedule will be affected by the value of the  $q_l^{\text{ref}}$ , which in turn is a function of the marginal cost  $\lambda$  and the total capacity of wells on the pad. That is, how much gas the wells can produce and the level of the peak-rates after shut-ins. To obtain a primal feasible solution and hence an upper bound  $Z^{\text{UB}}$ , we fix the binary variables  $y_{ljk}^d$  to the value obtained from solving the Lagrangian relaxation subproblems (20), and solve the primal problem (18). This variable fixing then corresponds to fixing the well schedule. The primal problem (18) reduces to an LP, where  $q_l^{\text{ref}}$  now are the only degrees of freedom. The LP solution of this variable fixing returns a set of primal feasible reference rates  $\bar{q}_l^{\text{ref}}$  which minimizes (in terms of the sum of the maximum norm) the deviation between the currently fixed production and the reference rates .

If the deviation between the primal feasible reference rates  $\bar{q}_l^{\text{ref}}$  from the binary variable fixing, and the corresponding solution  $q_l^{\text{ref}}$  from the Lagrangian relaxation is larger than a certain threshold,

$$|q_l^{\text{ref}} - \bar{q}_l^{\text{ref}}| > \delta^{\text{ref}}, \quad (29)$$

we perform a local search for each of these pads  $l \in \bar{\mathcal{L}} \subseteq \mathcal{L}$  to search for an improved shut-in schedule. Let  $\bar{Z}_l = \bar{f}_l$  be the solution from (18a) for pad  $l$  obtained by the binary variable fixing. The local search is then performed by fixing the reference rate,  $q_l^{\text{ref}} \equiv \bar{q}_l^{\text{ref}}$ , and solving the MILP problem

$$Z^l = \min f_l \quad (30a)$$

s.t.

$$f_l \leq \bar{Z}^l \quad (30b)$$

$$f_l \geq \bar{q}_l^{\text{ref}} - q_{lk}^{\text{pad}}, \quad k \in \mathcal{K} \quad (30c)$$

$$f_l \geq -\bar{q}_l^{\text{ref}} + q_{lk}^{\text{pad}}, \quad k \in \mathcal{K} \quad (30d)$$

$$\sum_{(l,j,k) \in \{(\bar{\mathcal{L}} \times \mathcal{J} \times \mathcal{K}) : \bar{y}_{ljk}^1 = 1\}} (1 - y_{ljk}^1) + \sum_{(l,j,k) \in \{(\bar{\mathcal{L}} \times \mathcal{J} \times \mathcal{K}) : \bar{y}_{ljk}^2 = 1\}} (1 - y_{ljk}^2) \leq r, \quad (30e)$$

eq. (18e)–(18w) for given  $l \in \bar{\mathcal{L}}$ .

The constraint (30e) is a local branching type constraint [52], where  $\bar{y}_{ljk}^1, \bar{y}_{ljk}^2$  is the shut-in schedule obtained from the solving the Lagrangian relaxation, and the parameter  $r$  is a neighborhood radius (or Hamming distance). This constraint allows at most  $r$  switchings of wells from on to off and from off to on, given by the first and the second term, respectively. The radius  $r$  must be selected sufficiently large for improved solutions to exist, while at the same time being small enough such that the local search can be performed numerically efficient. We implement a strategy where we initialize  $r = \lceil (1/3) |\mathcal{K}| |\mathcal{J}| \rceil$ , and increase  $r$  by 50% a maximum of five times if (30) is infeasible, and we allocate a maximum CPU time between 15-90 seconds for each local search, depending on the problem size.

If the loop in Lagrangian scheme terminates with a nonzero duality gap, we perform an extended but similar local search like (30) for the *best found* feasible solution. In particular, we remove the local branching constraint (30e) and perform a time-limited search, with  $q_l^{\text{ref}}$  now fixed to the best found primal feasible solution. In contrast, the local search (30) inside the loop is based on the deviation between the currently obtained feasible solution and solution from the current iteration of the Lagrangian relaxation. Note that we preserve primal feasibility when performing the local search. If the solution improves the upper bound, we update the best feasible solution and terminates the scheme.

The complete Lagrangian relaxation scheme is outlined in Figure 7.

## 6. Computational Results

The performance of the Lagrangian relaxation scheme is assessed by solving a series of test sets of the shale-gas multi-pad system with the structure shown in Figure 1. The sets are generated with different number of pads  $|\mathcal{L}|$ , while we for simplicity have used six wells per pad  $l$ . The proxy models in the numerical examples are realized and tuned against the multi-fracture reference model as described in section 3, using three different permeabilities  $k_m \in [1 \times 10^{-4}, 3 \times 10^{-4}, 5 \times 10^{-4}]$ . When using equal fracture geometry and initial pressure, the value of the formation permeability is decisive for the time when the gas rate first crosses the critical rate  $q_{gc}$ , hence giving somewhat different performance of the wells. We assume equal permeability for each well on given pad, but we will assume that each well have been operated

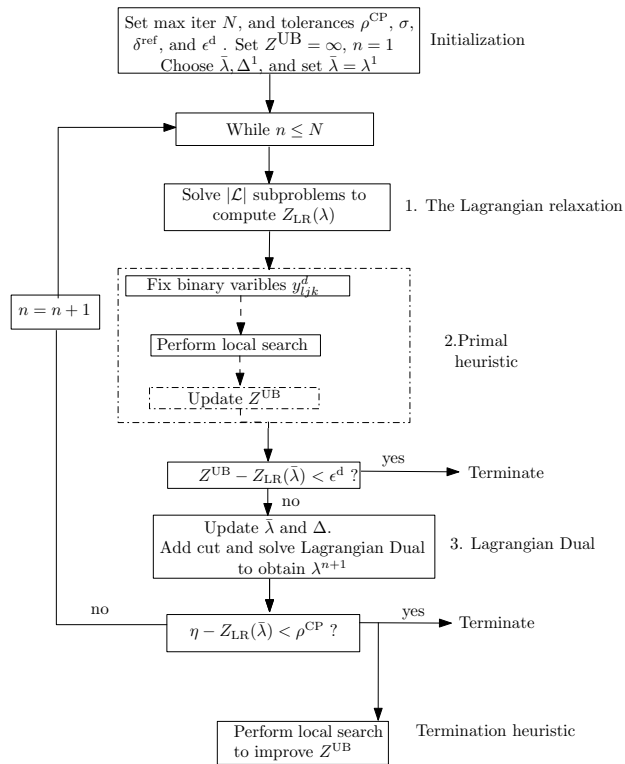


Figure 7: Schematic description of Lagrangian relaxation scheme.

different lengths of time, hence giving different initial pseudopressures  $m_{I_j}^{\text{init}}$ . We use a one-day fixed timestep, and set  $\tau_1 = \tau_2 = 2$  days.

The entire Lagrangian scheme illustrated in Figure 7 is implemented in GAMS [30], using IBM CPLEX v12.3 to solve the LPs and the MILPs. The problems are solved using a Dell laptop with Intel I7 quad-core CPU and 8GB of RAM, using deterministic parallel mode with up to 8 threads. The tolerances are set to  $\rho^{\text{CP}} = 10^{-6}$ ,  $\sigma = 0.01$ ,  $\delta^{\text{ref}} = 0.01$  and  $\epsilon^{\text{d}} = 0.01$ , while the maximum number of iteration are set to  $N = 15$ . The trust-region is initially set to  $\Delta_1 = 0.3$  and  $\bar{\lambda} = -0.5$ . For the MILP Lagrangian relaxation subproblems (20), we use a relative gap less than 1% or an absolute gap less than  $10^{-4}$  as stopping criteria, and allocate a maximum CPU time of four hours for each of the subproblems. The absolute-gap criteria is added as a the LP relaxation for the subproblems tends to be zero or a very small value, since we are minimizing the deviation between a reference and the rate. The size of the problems are shown in Table 3. Note that the size of the MILPs are significantly reduced by the presolve routines in CPLEX. All reported times are real (clock) time, and the reported gaps are defined as [42]

$$\text{DG} := 100[\%] \times \frac{|Z^{\text{UB}} - Z_{\text{LB}}|}{|Z^{\text{UB}}|}. \quad (31)$$

The results of the Lagrangian scheme is compared in Table 4 with the fullspace solution of the primal MILP model (18). We allocate a solution time for the fullspace solution approximately equal to the time spent by the Lagrangian scheme.

Table 3: Problem parameters before reduction in presolve routines

		Problem size		
$ \mathcal{L} $	$K$	Binary var.	Cont. var	Constraints
3	7	720	895	2666
3	14	1350	1672	5103
3	21	1980	2449	7580
6	7	1140	1789	5330
6	14	2700	3343	10244
6	21	3960	4897	15158
9	7	2160	2686	7994
9	14	4050	5014	15365
9	21	5940	7345	22736

Table 4 shows characteristics of the results of applying the Lagrangian relaxation scheme to sets with different number of pads and planning horizons. All of the sets in Table 4, except for the sets with  $|\mathcal{L}| = 6$  and  $K = 14$  and 21, respectively, are terminated by the criteria  $\tilde{\rho} \leq \rho^{\text{CP}}$  in step 2 of the trust-region method for solving the Lagrangian Dual (21). None of the sets reach the maximum number of iterations. Several of the sets terminate with duality gaps greater than 10%. Hence, if only evaluated by size of the duality gap, the performance of the Lagrangian relaxation scheme applied to these sets may be considered as only average. The final duality gaps are not necessarily worse for longer compared to short planning horizons  $K$ . For planning horizons longer than three weeks, the time required to solve each of the MILP subproblems to global optimality becomes computationally prohibitive. The difficulty in solving the MILP (18) becomes evident when comparing the solution characteristics of the Lagrangian scheme with solving the fullspace problem (18) with a global method, that is, a direct approach. The

Table 4: Computational results of Lagrangian relaxation scheme and a direct fullspace approach.

$\mathcal{L}$	$K$	$q^{\text{REF}}$ [ $10^5 \text{ m}^3/\text{d}$ ]	Lagrangian scheme						Fullspace			
			DG [%]	Time [min]	#iter	$Z^{\text{UB}}$	$Z_{\text{LR}}^{\ddagger}$	DG [%]	Time [min]	$Z^{\text{UB}}$	Best $Z_{\text{LB}}$	
3	7	1.6	22.9	1	7	0.41	0.32	100	2	0.50	0	
3	14	1.6	11.9	20	6	0.49	0.43	100	20	1.60	0	
3	21	1.6	11.7	151	5	0.52	0.46	100	84*	4.37	0	
6	7	4.3	2.4	2	7	2.09	2.04	90.4	2	2.57	0.25	
6	14	4.3	0.0	33	3	3.59	3.59	64.6	33	4.30	1.52	
6	21	4.3	0.0	580	5	4.13	4.13	63.3	580	6.67	2.45	
9	7	5.3	5.4	3	7	0.91	0.86	100	3	4.83	0	
9	14	5.3	14.6	43	9	1.55	1.32	100	43	9.42	0	
9	21	5.3	13.4	643	9	1.73	1.50	100	107*	18.1	0	

$\ddagger$  Results  $Z$  are deviations in the total rate, reported as  $10^4 \text{ m}^3/\text{d}$ .

\* Fullspace problems eventually ran out of memory due to size of nodefile in the order  $\sim 10\text{GB}$

Lagrangian relaxation scheme consistently finds significantly improved lower and upper bounds compared to the global method. Solving (18) by a global (branch-and-cut) method also highlights the difficulty in improving the lower bound; six out of nine sets in Table 4 are terminated with a 100% duality gap when solved with the direct approach, meaning that the best lower bound remains close to zero (cf. (31)) throughout the enumeration.

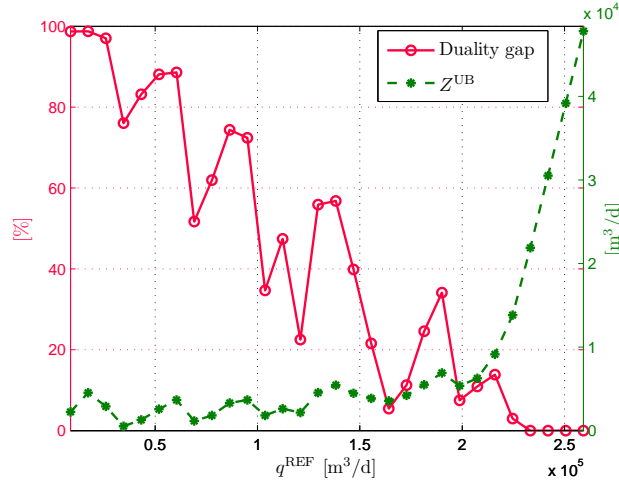


Figure 8: The duality gap and the best primal feasible solution (i.e. sum of deviations from  $q_i^{\text{ref}}$ ) as a function of the given total reference rate  $q^{\text{REF}}$  for the first row in Table 4.

The duality gap by termination of the Lagrangian relaxation scheme depends on the given reference rate  $q^{\text{REF}}$ . This is shown in Figure 8 for the set consisting of three pads and a one week planning horizon. For low values of  $q^{\text{REF}}$ , the gas-rate capacity of the multi-pad system is sufficient to meet the reference rate, seen by the low value of  $Z^{\text{UB}}$ , and hence a low total deviation between the produced pad rates and the corresponding reference rates. The duality gap, however, is large for low values of  $q^{\text{REF}}$ , and decreases for higher values of  $q^{\text{REF}}$ . In a



certain range of values for  $q^{\text{REF}}$ , it is seen that the size the duality gap compared the value of  $Z^{\text{UB}}$  is balanced. For high values of  $q^{\text{REF}}$ , the duality gap converges to zero while the optimal solution value increases rapidly, leading to unacceptable deviations between the produced rates and the reference rates.

The trust-region cutting-plane algorithm works well for solving the Lagrangian Dual of the problem. All sets terminate with an optimal dual variable  $\bar{\lambda}$  after 4-9 iterations. The method is also observed to be fairly robust with respect to initial choices of  $\bar{\lambda}$  and  $\Delta^i$ . Still, choosing a higher value of the initial trust region and a lower initial value of  $\bar{\lambda}$  (closer to -1) leads to slightly more iterations and to increased oscillations in the sequence of solutions  $\lambda^n$ . In this case the trust-region becomes ineffective. The use of the local search procedure (30) mainly improved the upper bound  $Z^{\text{UB}}$  whenever the binary-fixing heuristic found an improved primal feasible solution. In these cases, however, the local search often reduces the upper bound with 50-200%, relatively.

The practical performance of the solving (18) is illustrated in Figure 9 showing the total rate for the set in Table 4 with 9 wells and a three week planning horizon. The reference rate and the total produced rate is shown for three of the nine pads in Figure 10. The total gas rate for the pads is shown to tightly follow the given total reference rate  $q^{\text{REF}}$  over the planning horizon, while each of the pads rates follow the optimal pad reference rates with only small fluctuations as seen in Figure 10. Consequently, any high and low peak rates are avoided, both in the total rate, and for the rates from the individual pads. In this context, the actual optimized result by the Lagrangian relaxation scheme clearly renders a good feasible solution despite the duality gap being more than 13%. Comparing the subfigures in Figure 10 also highlights a qualitative difference in the performance of the gas rates  $q_{lk}^{\text{pad}}$  from the pads; the wells with the highest permeability preserve a higher rate after shut-ins. Hence, fewer shut-ins are needed, and the total rate from the pad is more stable as seen in Figure 10c.

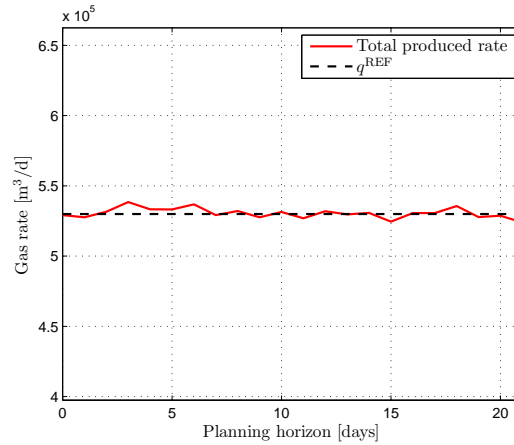


Figure 9: Total rate from  $|\mathcal{L}| = 9$  pads with a three week planning horizon solved with the Lagrangian relaxation scheme.

## 7. Discussion

The main contribution in this paper is the development of a decomposable Lagrangian relaxation scheme from a fullspace GDP formulation for field-wide shut-in scheduling in shale-gas

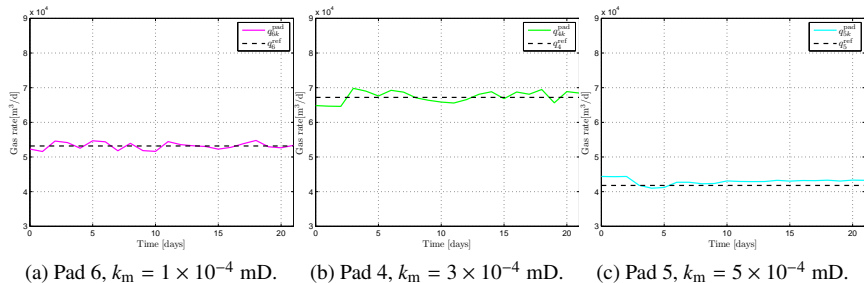


Figure 10: Pad reference rates and produced rates.

systems. The scheme was observed to consistently solve the Lagrangian dual and find feasible solutions through a heuristic using the solutions from the Lagrangian relaxation. The Lagrangian scheme clearly outperformed a direct approach, both in terms of the duality gap and the lower and upper bounds. The schemes' ability to generate small duality gaps, however, is observed to depend somewhat on the problem configuration. Nevertheless, the solutions obtained are still considered as good feasible solutions. If, however, very small duality gaps are desirable (or required), then using either Augmented Lagrangian [48] with a decomposition strategy similar to [53] can be considered, or integrating the Lagrangian relaxation (19) in a branch-and-bound enumeration scheme (i.e., a branch-and-price framework).

The computational challenges experienced with improving the lower bound is related to two aspects of the problem formulation. The first is the strength of the lower bound provided by the Lagrangian relaxation, which may depend strongly on the chosen relaxation. If the variables in the linking constraints also appear in several of the other constraints in the subproblems, then applying Lagrangian *decomposition* [54] is generally preferable, in the sense that the lower bound in Lagrangian decomposition is as least as tight as any other Lagrangian relaxation. The complexity, on the other hand, in solving the scheduling problem for the shale-gas multi-pad system is inherent, as the problem involves both dynamics, combinatorial decisions and large-scale systems. The scheduling is based on switching wells on-off to track a variable reference rate. As we assume that each well has a separate independent closed-loop control of the wellhead pressure, then meeting the reference rate *exactly* will almost certainly be impossible.

A possible extension of the proposed Lagrangian relaxation scheme is to consider decomposition per wells, that is, decomposition over the index  $\mathcal{J}$ . A viable approach may then be a bi-level decomposition scheme, in which the current Lagrangian relaxation is retained while the subproblems (20) are further decomposed into one problem per well. However, the reformulation (18c)–(18d) of the maximum norm can in this case no longer be applied. Hence, either a different objective function or a special purpose algorithm for  $\ell_1$  or  $\ell_\infty$ -type problems must be applied. A temporal decomposition of the primal MILP (18) over the index  $\mathcal{K}$  may also be considered, however, leading to an enormous increase in the number of dual variables. Note that a simple acceleration scheme for the given Lagrangian relaxation scheme is parallelization of the subproblems using grid-computing.

The Lagrangian relaxation (19) of the primal problem (18) leads to a decomposition with subproblems that are similar in size and structure as the formulation of the single-pad tracking problem in [10]. The subproblems (20), however, include variable  $q_i^{\text{ref}}$ , leading to a problem which are considerably harder to solve. The scheme can easily be extended to handle a varying

reference rate  $q_k^{\text{REF}}$ . This is illustrated in Figure (11) for a system of three pads and a three week planning horizon. Still, the Lagrangian relaxation scheme becomes computationally more difficult with varying reference rate, and gives a linear increase in the number of dual variables. Consequently, if the number of dual variables becomes close to the planning horizon, the cutting-plane method should be stabilized by a quadratic term in the objective instead of the trust-region scheme in (25), see [48].

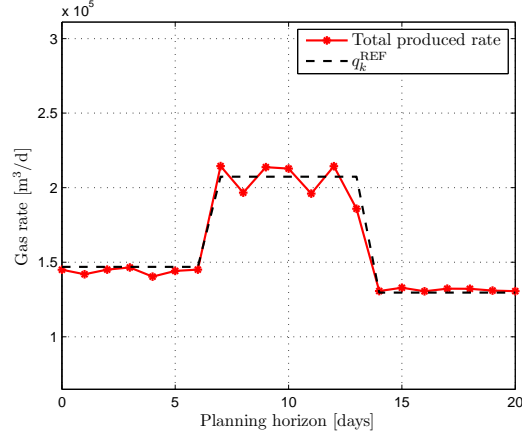


Figure 11: Example on time-varying given reference rate  $q_k^{\text{REF}}$

The paper also includes a systematic, frequency selective tuning procedure for the shale well proxy model, together with a simple scheme for updating the parameters in the model. By properly designing the filter for the prediction errors, it is possible to select which of the dynamics in shale-gas wells the proxy model should emphasize. The simplicity of the proxy model still limits to some extent the range of dynamics the model can cover. The tuning procedure further consists of solving a nonconvex NLP, possibly leading to suboptimal parameters in the proxy model. This can be compensated for by frequently updating the model parameters solving the MHE type estimation problem (10).

Scheduling shut-ins of shale-gas systems relative to a reference rate rather than maximizing the production may be beneficial both from an operational as well as an economic and energy perspective. Operationally it leads to more stable rates as seen in Figures 9 and 10, with the potential of significantly reducing high and low peaks compared to a standard approach with fixed shut-in times, see [10]. The economic reason is related to the current low sales prices of natural gas a result of increased availability, mainly due to overproduction of shale gas. Gas prices may vary substantially with seasonal demands, and produced natural gas may be stored and sold at times with higher prices. As mentioned earlier, the loss in cumulative production by performing short shut-ins may be kept very low for shale-gas wells due to the fracture networks acting as a recharge and storage system. Consequently, by optimizing well shut-ins with respect to varying demands as illustrated in Figure 11, the wells may be scheduled to account for seasonal variations in demands and gas prices. The high availability of shale-gas, together with the shut-in characteristics and properties described and optimized in this paper, makes it possible to utilize shale-gas systems as a buffer of natural gas in an energy and power planning context.

## 8. Conclusions

In this work, a novel Lagrangian relaxation based scheme for shut-in scheduling in shale-gas multi-pad systems has been presented. The shut-in scheduling is integrated in field-wide production planning based on demands in gas-rate, and it is shown how production objectives and constraints at different levels in the large-scale system can be included in the formulation. The decomposable scheme is scalable in the number of pads and shown through computational testing to outperform a direct approach for all sizable problems, though the efficiency of the scheme is observed to decrease for long planning horizons.

## 9. Acknowledgments

This work was supported by the Center for Integrated Operations in the Petroleum Industry, Trondheim, Norway (IO center). The authors gratefully acknowledge valuable feedback from Professor Curtis H. Whitson at NTNU, and financial support from an IBM PhD scholarship.

## References

- [1] Annual Energy Outlook 2011, Tech. Rep., The U.S. Energy Information Administration, 2011.
- [2] Z. Dong, S. A. Holditch, D. A. Mcvay, W. B. Ayers, Global Unconventional Gas Resource Assessment, *SPE Economics & Management* 4 (4) (2012) 222–234.
- [3] R. Stefik, K. Paulson, When Unconventional Becomes Conventional, *Journal of Canadian Petroleum Technology* 50 (11) (2011) 68–70.
- [4] C. D. Jenkins, C. M Boyer II, Coalbed- and Shale-Gas Reservoirs, *Journal of Petroleum Technology* 60 (2) (2008) 92–99.
- [5] J. Baihly, R. Altman, R. Malpani, F. Luo, Shale Gas Production Decline Trend Comparison Over Time and Basins, in: *SPE Annual Technical Conference and Exhibition*, Florence, Italy, sPE 135555, 2010.
- [6] C. H. Whitson, S. D. Rahmawati, A. Juell, Cyclic Shut-in Eliminates Liquid-Loading in Gas Wells, in: *SPE/EAGE European Unconventional Resources Conference and Exhibition*, March, Vienna, Austria, SPE 153073, 2012.
- [7] H. Al Ahmadi, A. Almarzooq, R. Wattenbarger, Application of Linear Flow Analysis to Shale Gas Wells - Field Cases, in: *SPE Unconventional Gas Conference*, Pittsburgh, Pennsylvania, USA, SPE 130370-MS, 2010.
- [8] J. F. Lea, H. V. Nickens, Solving Gas-Well Liquid-Loading Problems, *Journal of Petroleum Technology* 56 (4) (2004) 30–36, SPE 72092.
- [9] N. Dousi, C. Veeken, P. K. Currie, Numerical and Analytical Modeling of the Gas-Well Liquid-Loading Process, *SPE Production & Operations* 21 (4) (2006) 6–9.
- [10] B. R. Knudsen, B. Foss, Shut-in Based Production Optimization of Shale-gas Systems, *Computers and Chemical Engineering* doi:10.1016/j.compchemeng.2013.05.022.
- [11] B. Knudsen, B. Foss, C. Whitson, A. Conn, Target-rate Tracking for Shale-gas Multi-well Pads by Scheduled Shut-ins, in: *IFAC Proceedings of the International Symposium on Advanced Control of Chemical Processes*, Singapore, 107–113, 2012.
- [12] B. Foss, V. Gunnerud, M. Dueñas Díez, Lagrangian Decomposition of Oil-Production Optimization Applied to the Troll West Oil Rim, *SPE Journal* 14 (4).
- [13] C. Sagastizábal, Divide to conquer: decomposition methods for energy optimization, *Mathematical Programming* 134 (1) (2012) 187–222.
- [14] J. Warren, P. Root, The Behavior of Naturally Fractured Reservoirs, *SPE Journal* 3 (3) (1963) 245–255.
- [15] M. Karimi-Farad, L. J. Durlofsky, K. Aziz, An Efficient Discrete-Fracture Model Applicable for General-Purpose Reservoir Simulators, *SPE Journal* 9 (2) (2004) 227–236.
- [16] C. L. Cipolla, E. Lolon, J. Erdle, B. Rubin, Reservoir Modeling in Shale-Gas Reservoirs, *SPE Reservoir Evaluation & Engineering* 13 (4) (2010) 638–653.
- [17] R. P. Sutton, S. A. Cox, R. D. Barree, Shale Gas Plays : A Performance Perspective, in: *SPE Tight Gas Completions Conference*, San Antonio, Texas, USA, sPE 138447-MS, 2010.
- [18] R. O. Bello, R. A. Wattenbarger, Modelling and Analysis of Shale Gas Production With a Skin Effect, *Journal of Canadian Petroleum Technology* 49 (12) (2010) 37–48.

- [19] R. Al-Hussainy, H. R. Jr., P. Crawford, The Flow of Real Gases Through Porous Media, *Journal of Petroleum Technology* 18 (5) (1966) 624–636.
- [20] C. H. Whitson, M. R. Brule, Phase Behavior, vol. 20 of *SPE Monograph Series*, Society of Petroleum Engineers, 2000.
- [21] R. Turner, M. Hubbard, A. Dukler, Analysis and Prediction of Minimum Flow Rate for the Continuous Removal of Liquids from Gas Wells, *Journal of Petroleum Technology* 21 (11) (1969) 1475–1482.
- [22] S. B. Coleman, H. B. Clay, D. G. Mccurdy, H. L. Norris III, A New Look at Predicting Gas-Well Load-Up, *Journal of Petroleum Technology* 43 (3) (1991) 329–333.
- [23] R. Jayakumar, V. Sahai, A. Boulis, A Better Understanding of Finite Element Simulation for Shale Gas Reservoirs through a Series of Different Case Histories, in: *SPE Middle East Unconventional Gas Conference and Exhibition*, Muscat, Oman, SPE 142464, 2011.
- [24] L. Ljung, *System Identification: Theory for the user*, Prentice Hall, second edn., 1999.
- [25] W. Spinelli, L. Piroddi, M. Lovera, On the role of prefiltering in nonlinear system identification, *IEEE Transactions on Automatic Control* 50 (10) (2005) 1597–1602.
- [26] M. L. Thompson, M. A. Kramer, Modeling chemical processes using prior knowledge and neural networks, *AIChE Journal* 40 (8) (1994) 1328–1340.
- [27] T. A. Johansen, Identification of non-linear systems using empirical data and prior knowledgean optimization approach, *Automatica* 32 (3) (1996) 337–356.
- [28] *SENSOR Manual*, Coats Engineering Inc., 2009.
- [29] D. L. Katz, R. L. Lee, *Natural Gas Engineering*, McGraw-Hill Publishing Company, 1990.
- [30] A. Brooke, D. Kendrick, A. Meeraus, R. Raman, *GAMS - A user's guide*, 2011.
- [31] B. Foss, J. P. Jensen, Performance Analysis for Closed-Loop Reservoir Management, *SPE Journal* 16 (1) (2011) 183–190, doi:10.2118/138891-PA.
- [32] H. P. Bieker, O. Slupphaug, T. A. Johansen, Real-Time Production Optimization of Oil and Gas Production Systems : A Technology Survey, *SPE Production & Operations* 22 (4) (2007) 382–391.
- [33] C. V. Rao, J. B. Rawlings, J. H. Lee, Constrained linear state estimation - a moving horizon approach, *Automatica* 37 (10) (2001) 1619–1628.
- [34] R. Raman, I. E. Grossmann, Modelling and computational techniques for logic based integer programming, *Computers & Chemical Engineering* 18 (7) (1994) 563–578.
- [35] I. E. Grossmann, F. Trespalcios, Systematic Modeling of Discrete-Continuous Optimization Models through Generalized Disjunctive Programming, *AIChE Journal* doi:10.1002/aic.
- [36] G. L. Nemhauser, L. A. Wolsey, *Integer and Combinatorial Optimization*, John Wiley & Sons, Inc., 1988.
- [37] A. Vecchiotti, I. E. Grossmann, Modeling issues and implementation of language for disjunctive programming, *Computers & Chemical Engineering* 24 (2000) 2143–2155.
- [38] E. Balas, Disjunctive programming and a hierarchy of relaxations for discret optimization problems., *SIAM Journal on Algebraic and Discrete Methods* 6 (3) (1985) 466–486.
- [39] A. Vecchiotti, S. Lee, I. E. Grossmann, Modeling of discrete/continuous optimization problems: characterization and formulation of disjunctions and their relaxations, *Computers & Chemical Engineering* 27 (3) (2003) 433–448.
- [40] R. Raman, I. Grossmann, Relation between MILP modelling and logical inference for chemical process synthesis, *Computers & Chemical Engineering* 15 (2) (1991) 73–84.
- [41] S. Takriti, B. Krasenbrink, L. S.-Y. Wu, Incorporating fuel constraints and electricity spot prices into the stochastic unit commitment problem, *Operations Research* 48 (2) (2000) 268–280.
- [42] Y. Pochet, L. A. Wolsey, *Production Planning by Mixed Integer Programming*, Springer, 2006.
- [43] J. Lee, J. Leung, F. Margot, Min-up/min-down polytopes, *Discrete Optimization* 1 (1) (2004) 77–85.
- [44] D. Rajan, S. Takriti, Minimum Up / Down Polytopes of the Unit Commitment Problem with Start-Up Costs, Tech. Rep., IBM Research Report, 2005.
- [45] A. M. Geoffrion, Lagrangian Relaxation and its Uses in Integer Programming, *Mathematical Programming Study* 2 (1974) 82–114.
- [46] A. Frangioni, About Lagrangian Methods in Integer Optimization, *Annals of Operations Research* 139 (1) (2005) 163–193.
- [47] J. E. Kelley, The Cutting-Plane Method for Solving Convex Programs, *Journal of the Society for Industrial and Applied Mathematics* 8 (4) (1960) 703–712.
- [48] J.-B. Hiriart-Urruty, C. Lemarechal, *Convex Analysis and Minimization Algorithms: part II*, Springer, 1993.
- [49] S. Mouret, I. E. Grossmann, P. Pestaiaux, A new Lagrangian decomposition approach applied to the integration of refinery planning and crude-oil scheduling, *Computers & Chemical Engineering* 35 (12) (2011) 2750–2766.
- [50] B. Kallehauge, J. Larsen, O. B. Madsen, Lagrangian duality applied to the vehicle routing problem with time windows, *Computers & Operations Research* 33 (5) (2006) 1464–1487.
- [51] R. E. Marsten, W. W. Hogan, J. W. Blankenship, The Boxstep Method for Large-Scale Optimization, *Operations Research* 23 (2) (1975) 389–405.

- [52] M. Fischetti, A. Lodi, Local branching, *Mathematical Programming* 98 (1-3) (2003) 23–47.  
 [53] Z. Li, M. G. Ierapetritou, Capacity Expansion Planning Through Augmented Lagrangian Optimization and Scenario Decomposition, *AiChE Journal* 58 (3).  
 [54] M. Guignard, S. Kim, Lagrangean Decomposition: A Model Yielding Stronger Lagrangian Bounds, *Mathematical Programming* 39 (2) (1987) 215–228.  
 [55] R. K. Martin, *Large Scale Linear and Integer Optimization: A Unified Approach*, Kluwer Academic Publishers, 1998.

## Appendix A. Proof of bound on dual variable

For proving the bound on  $\lambda \geq -1$ , we can consider the following simplified Lagrangian relaxation subproblem (20):

$$\min f + \lambda q^{\text{ref}} \quad (\text{A.1a})$$

s.t.

$$f \geq q^{\text{ref}} - q^{\text{pad}}, \quad (\text{A.1b})$$

$$f \geq -q^{\text{ref}} + q^{\text{pad}}, \quad (\text{A.1c})$$

$$q^{\text{pad}} \leq \bar{U}, \quad (\text{A.1d})$$

$$f, q^{\text{ref}}, q^{\text{pad}} \geq 0, \quad (\text{A.1e})$$

where we for simplicity have omitted the indices  $k$  and  $l$ . The variables  $f, q^{\text{ref}}$  are block separable from the remaining variables defining the feasible set of (20), while the total rate  $q^{\text{pad}}$  is defined by a polytope  $\bar{Q}_2$  spanned by (18e)–(18w) for given  $l \in \mathcal{L}$ , i.e. a bounded polyhedron [36]. Consequently we can deduce an upper bound  $\bar{U}$  on  $q^{\text{pad}}$ . Consider the general LP,  $Z_{\text{LP}} = \min\{c^T x : Ax \geq b, x \geq 0\}$ , and the corresponding symmetric dual LP,  $Z_{\text{LPDual}} = \max\{b^T u : A^T u \leq c, u \geq 0\}$ . By weak duality, the primal LP is feasible if and only if its corresponding dual LP is feasible. Similarly, the primal LP is unbounded if the dual LP is infeasible [36]. Hence, by considering the feasible set  $A^T u \leq c, u \geq 0$  for the dual LP of (A.1) parametrized by  $\lambda$ , we obtain the linear inequalities,

$$u_1 + u_2 \leq 1, \quad (\text{A.2a})$$

$$-u_1 + u_2 \leq \lambda, \quad (\text{A.2b})$$

$$u_1 - u_2 - u_3 \leq 0, \quad (\text{A.2c})$$

$$-u_1 \leq 0, \quad (\text{A.2d})$$

$$-u_2 \leq 0, \quad (\text{A.2e})$$

$$-u_3 \leq 0. \quad (\text{A.2f})$$

Feasibility requirements of the above inequalities can be derived by Fourier-Motzkin elimination, see e.g. [55]. By successively eliminating  $u_1$  and  $u_2$  by pairing the inequalities (A.2), we obtain

the inequalities

$$-u_3 \leq \frac{1}{2}(1 + \lambda), \quad (\text{A.3a})$$

$$0 \leq \frac{1}{2}(1 + \lambda), \quad (\text{A.3b})$$

$$u_3 \leq 1, \quad (\text{A.3c})$$

$$-u_1 \leq 0, \quad (\text{A.3d})$$

$$0 \leq 1, \quad (\text{A.3e})$$

$$-u_3 \leq 0. \quad (\text{A.3f})$$

The bound  $\lambda \geq -1$  then immediately follows from (A.3b).