On the Solution of Ill-Posed Control and Estimation Problems: An Inertia-Free Approach

Victor M. Zavala
Department of Chemical and Biological Engineering
University of Wisconsin-Madison
1415 Engineering Hall, Madison WI 53706

Nai-Yuan Chiang*
Mathematics and Computer Science Division
Argonne National Laboratory
9700 S Cass Ave, Argonne, IL 60439

Abstract

We analyze the performance of an inertia-free, filter line-search approach for solving ill-posed nonlinear programs. Ill-posedness arises in applications such as economic model predictive control and estimation and is often the result of having many degrees of freedom, linear objective functions, and/or uninformative data. Ill-posedness manifests numerically as ill-conditioned reduced Hessian matrices with eigenvalues that are close to zero at intermediate iterates or stationary points. In the presence of ill-conditioning, inertia estimates obtained from factorization routines can become unreliable and might trigger unnecessary regularization of the Hessian, ultimately slowing down progress. We argue that an inertia-free approach is an effective alternative to handle such problems. We present a state estimation studies for weakly observable systems and a large-scale stochastic optimal control formulation with economic and quadratic tracking objectives to support this claim.

Keywords: inertia, large-scale, nonlinear optimization, ill-posed, control, estimation

*Present Address: United Technologies Research Center, 411 Silver Lane, East Hartford, CT, 06108

1 Introduction

Ill-posed optimization formulations are those in which stationarity points are non-isolated minimizers. In such problems, different variable combinations give the same optimal objective value. One can visualize this as an objective function that does not have curvature at a stationary point (e.g., it is flat). Ill-posedness can also manifest along the search, indicating that there is limited curvature at a particular iterate. Ill-posed problems arise in many domains but are particularly prevalent in parameter/state estimation formulations and economic model predictive control (MPC). Ill-posedness is observed in estimation settings, for instance, when the system states or parameters are not observable given the measured data. This is often corrected by adding regularization terms directly to the problem formulation in the form of Bayesian priors or Tikhnonov terms [23, 31, 2]. In the economic MPC case, ill-posedness arises when the control actions have limited influence on the cost function. This can also arise when the objective functions are linear (e.g., profit, throughput) and the constraints do not provide curvature [11]. Ill-posedness is also resolved in MPC by adding regularization terms to the cost function, often in the form of quadratic tracking terms [17]. Ill-posedness also arises quite naturally in large-scale applications where the number of degrees of freedom is large such as PDE-constrained optimization, stochastic optimization, and errors-in-variables-measured (EVM) estimation formulations [3, 33, 35, 26].

Ill-posed problems are difficult to deal with numerically. In the absence of curvature or on the presence of negative curvature, line-search based solvers need to regularize the problem on-the-fly in order to prevent singularity of the so-called KKT matrix and to ensure that the computed step is a descent direction. Regularization, however, decreases the quality of Newton steps and can slow down convergence (particularly near the solution). Regularization is performed in line-search implementations such as `IPOPT` by adding a positive definite matrix to the Hessian [28]. The absence of curvature or the presence of negative curvature is detected in modern optimization solvers by using inertia information of the KKT matrix. Such inertia information (number of positive, negative, and zero eigenvalues) is obtained from symmetric indefinite factorization codes such as `MA57` and `PARDISO` [22, 12, 25]. An important observation is that these inertia estimates are not necessarily dependent of the actual magnitude of the eigenvalues. For instance, if eigenvalues are negative, linear solvers might indicate that regularization is needed, even if the eigenvalues are small in magnitude. Moreover, inertia estimates will depend on the pivoting strategy and tolerance used. Because of this, inertia estimates can become unreliable in ill-posed problem formulations, ultimately resulting in decreased performance.

Trust-region strategies provide a more natural mechanism to deal with ill-posed problems. In particular, trust-region settings have the advantage that no explicit regularization of the Hessian is needed (this is done implicitly thought the trust-region constraint). Efficient handling of ill-posed formulations by trust-region algorithms has been demonstrated in the context of parameter estimation [1]. Modern trust-region implementations for constrained optimization decompose the search step into tangential and normal components. The tangential step is computed by approximately solving a trust-region subproblem [6, 14, 16]. This is typically done by using a projected conjugate gradient (PCG) scheme that detects negative curvature at the inner iterates. This is done by computing inner-

outer products with the Hessian on the null space of the Jacobian (instead of using inertia estimates as in the line-search approach). In the presence of a direction of negative curvature, the PCG path is continued along this direction until it reaches the trust-region boundary. This approach is guaranteed to be globally convergent because it can always improve the Cauchy step (or revert to it) [9]. Iter is well-known, however, that the quality of the step can be poor when the PCG procedure is terminated prematurely and this can result in excessive shrinking of the trust-region and slow progress.

Under a trust-region setting one is limited in the linear algebra techniques that can be used to compute the search step. In particular, one requires schemes that are compatible with the globalization approach (i.e., improve the progress of the Cauchy step). Linear algebra strategies such as direct factorizations, GMRES, and QMR are not fully compatible in this sense. This is important because some of these linear algebra strategies might handle difficult linear systems more efficiently than PCG. This observation has in fact motivated the implementation of a hybrid trust-region and line-search strategy in the widely used `KNITRO` package [29]. In addition, trust-region settings require tailored preconditioners that project (exactly or inexactly) the iterates onto the nullspace of the constraint Jacobian [14, 16]. This feature limits the types of preconditioners that can be used. A line-search setting has the practical advantage that any linear algebra scheme can be used to compute the step (as long as the computed step is a descent direction) and this provides computational flexibility. Such flexibility motivates our interest in line-search approaches. The price to pay in a line-search setting, however, is that inertia detection and regularization of the Hessian are necessary.

Motivated by curvature tests and the flexibility provided by trust-region settings, we have recently proposed inertia-free line-search strategies that perform simple curvature tests along the search directions [8, 7]. The curvature tests guarantee descent when the constraint violation is sufficiently small. When the test is not fulfilled, it triggers regularization of the Hessian matrix. Surprisingly enough, this simple approach enjoys the same global convergence of the standard inertia-based approach used in filter Iine-search settings such as that of `IPOPT`. We have also demonstrated that this approach can handle large-scale and highly nonlinear problems. In this work we analyze the performance of inertia-free approaches in ill-posed nonlinear programming (NLP) formulations. In particular, we present estimation formulations for weakly observable systems and a large-scale stochastic optimal control formulation for natural gas networks. We demonstrate that the inertia-free approaches can deal with ill-posedness effectively and present an attractive alternative to inertia-based settings.

The paper is structured as follows. In the Section 2 we present basic definitions and a standard inertia-based algorithmic setting. In Section 3 we present inertia-free settings and in Section 4 we present ill-posed estimation and control studies.

2 Basic Definitions and Setting

Consider the NLP of the general form

$$\min_{x \in \Re^n} \quad f(x) \tag{2.1a}$$

$$\text{s.t. } c(x) = 0 \tag{2.1b}$$

$$x \geq 0. \tag{2.1c}$$

Here, $x \in \Re^n$ are primal variables, and the objective and constraint functions are $f : \Re^n \to \Re$ and $c : \Re^n \to \Re^m$, respectively. We use a logarithmic barrier framework with subproblems of the form

$$\min_{x \in \Re^n} \varphi^\mu(x) := f(x) - \mu \sum_{j=1}^{n} \ln x^{(j)} \tag{2.2a}$$

$$\text{s.t. } c(x) = 0 \tag{2.2b}$$

where $\mu > 0$ is the barrier parameter and $x^{(j)}$ is the jth entry of vector x. We consider a framework that solves a sequence of barrier problems (2.2) and drives the barrier parameter μ to zero. To approximately solve each barrier problem, we apply Newton's method to its optimality conditions:

$$\nabla_x \varphi^\mu(x) + \nabla_x c(x)\lambda = 0 \tag{2.3a}$$

$$c(x) = 0 \tag{2.3b}$$

while enforcing $x \geq 0$ along the search. Here, $\lambda \in \Re^m$ are multipliers for equality constraints. The primal variables and multipliers at iteration k are denoted as (x_k, λ_k). Their corresponding search directions $(d_k, \lambda_k^+ - \lambda_k)$ can be obtained by solving the linear system

$$\begin{bmatrix} W_k & J_k^T \\ J_k & 0 \end{bmatrix} \begin{bmatrix} d_k \\ \lambda_k^+ \end{bmatrix} = - \begin{bmatrix} g_k \\ c_k \end{bmatrix}. \tag{2.4}$$

We refer to this system as the augmented system. Here, λ_k^+ is the multiplier update, $c_k := c(x_k)$, $J_k := \nabla_x c(x_k)^T \in \Re^{m \times n}$, $g_k := \nabla_x \varphi_k^\mu$, $W_k := H_k + \Sigma_k$, $H_k := \nabla_{xx} \mathcal{L}(x_k, \lambda_k) \in \Re^{n \times n}$, $\mathcal{L}(x_k, \lambda_k) := f(x_k) + \lambda_k^T c(x_k)$, and $\Sigma_k := X_k^{-2}$ with $X_k := \text{diag}(x_k)$. One can show that the primal-dual approximation $\Sigma_k \approx X_k^{-1} V_k$, where $V_k := \text{diag}(\nu_k)$ and ν_k are multiplier estimates for the bounds (2.1c), can be used as long as the products $x_k^{(j)} \nu_k^{(j)}$ remain proportional to μ [27, 10]. To enable compact notation, we define the augmented matrix

$$M_k := \begin{bmatrix} W_k & J_k^T \\ J_k & 0 \end{bmatrix}. \tag{2.5}$$

We can also consider the computation of the search directions d_k using the decomposition

$$d_k = n_k + t_k. \tag{2.6}$$

Here, n_k is computed from

$$\begin{bmatrix} W_k & J_k^T \\ J_k & 0 \end{bmatrix} \begin{bmatrix} n_k \\ \cdot \end{bmatrix} = - \begin{bmatrix} 0 \\ c_k \end{bmatrix}, \tag{2.7}$$

and t_k is computed from

$$\begin{bmatrix} W_k & J_k^T \\ J_k & 0 \end{bmatrix} \begin{bmatrix} t_k \\ \lambda_k^+ \end{bmatrix} = - \begin{bmatrix} g_k + W_k n_k \\ 0 \end{bmatrix}. \tag{2.8}$$

The inertia of the augmented matrix M_k is defined as its number of positive ℓ_+, negative ℓ_-, and zero ℓ_0 eigenvalues. We use the standard notation $\text{Inertia}(M_k) = \{\ell_+, \ell_-, \ell_0\}$. One can show that, if the Jacobian has full row rank, the following identity holds:

$$\text{Inertia}(M_k) = \text{Inertia}(Z_k^T W_k Z_k) + \{m, m, 0\} \tag{2.9}$$

where matrix Z_k lies on the null space of the Jacobian J_k (i.e., $J_k Z_k = 0$) and $Z_k^T W_k Z_k$ is the reduced Hessian. The reduced Hessian is an $(n-m) \times (n-m)$ matrix, where $n-m$ is the number of degrees of freedom. Positive definiteness of the reduced Hessian (i.e., $\text{Inertia}(Z_k^T W_k Z_k) = \{n-m, 0, 0\}$) guarantees that the search direction is of descent when the constraint violation is sufficiently small (see Lemma 2 in [27]). This descent property is essential for establishing global convergence in the filter line-search setting that we consider here [8].

Positive definiteness of the reduced Hessian can be enforced in a practical setting by monitoring the inertia of the augmented matrix M_k and correcting it (if necessary) by regularizing the Hessian matrix as $W_k \leftarrow W_k + \delta I$ for $\delta \geq 0$. This *convexification* approach is justified from the relationship (2.9) which shows that the reduced Hessian is positive definite if and only if the augmented matrix M_k has n positive, m negative, and no zero eigenvalues. We state this inertia condition formally as

$$\text{Inertia}(M_k) = \{n, m, 0\}. \tag{2.10}$$

The augmented matrix M_k can be factored as LBL^T by using symmetric indefinite factorizations where L is a unit lower triangular matrix and B is a block diagonal matrix composed of 1×1 and 2×2 diagonal blocks. From Sylvester's law of inertia we know that the eigenvalues of M_k are the eigenvalues of B. Furthermore, because each 2×2 block is constructed by having one positive and one negative eigenvalue, the inertia of M_k can be estimated from the inertia of B [5].

The following procedure is often used to regularize the Hessian matrix. The matrix M_k is decomposed as LBL^T for $\delta = 0$ and the search direction d_k is computed using this decomposition. If the inertia is correct (i.e., condition (2.10) is satisfied), the search direction d_k is used as trial step in the filter line-search procedure. If the inertia is not correct, the regularization parameter δ is increased, the augmented matrix is refactorized and a new direction d_k is obtained. The procedure is repeated until the matrix M_k has the correct inertia. Heuristics are incorporated to accelerate the rate of increase/decrease of δ in order to ensure that the number of trial factorizations is not too large (because each factorization is computationally expensive). Moreover, we note that large regularization can deteriorate the quality of the Newton step and slow down convergence. The inertia-based regularization strategy implemented in the current version of `IPOPT` [28] is shown below. Here, $0 < \bar{\delta}^{min} < \bar{\delta}^0$, $0 < \kappa^- < 1 < \kappa^+ < \hat{\kappa}^+$ are given constants.

INERTIA-BASED REGULARIZATION (IBR)

IBR-1 Factorize M_k with $\delta = 0$. If (2.10) holds, compute d_k and stop.

IBR-2 If $\delta^{last} = 0$, set $\delta \leftarrow \bar{\delta}^0$, otherwise set $\delta \leftarrow \max\{\bar{\delta}^{min}, \kappa^- \delta^{last}\}$.

IBR-3 Factorize M_k with current δ. If (2.10) holds, set $\delta^{last} \leftarrow \delta$, compute d_k and stop.

IBR-4 If $\delta^{last} = 0$, set $\delta \leftarrow \hat{\kappa}^+ \delta$, otherwise set $\delta \leftarrow \kappa^+ \delta$ and go to IBR-3.

We make the following remarks:

- When the problem is ill-posed some eigenvalues of the reduced Hessian will be close to zero with some of them possibly being slightly negative. This can make the inertia estimates of the LBL^T factorization unreliable (i.e., sensitive to numerical errors) and thus the inertia test might trigger regularization unnecessarily and slow down progress. When the eigenvalues are sufficiently away from zero, the reduced Hessian might have large negative eigenvalues, which implies that the problem is locally at a saddle point. This raises an interesting observation: consider the case in which an NLP does not have an isolated local minimum (e.g., only has a saddle point or stationarity point with zero curvature). In such a case, the inertia-based approach will try to regularize at each iteration and slow progress will be observed.

- When the cost function $f(\cdot)$ and the constraints $c(\cdot)$ are linear, we have that $W_k = \Sigma_k$ and the augmented matrix has the structure:

$$M_k := \begin{bmatrix} \Sigma_k & J_k^T \\ J_k & 0 \end{bmatrix}. \tag{2.11}$$

  Consequently, the only curvature seen is through the primal-dual term Σ_k corresponding to the bound constraints. This explains why the use of linear objectives might introduce ill-conditioning. This is particularly relevant in economic MPC formulations.

- The addition of quadratic regularization terms to the objective function (e.g., $f(x) + \kappa\|x - \bar{x}\|^2$ for some $\kappa > 0$) has a similar effect to that of correcting the inertia by adding the term $\sigma\mathbb{I}$ to the Hessian. Inertia correction, however, is adaptive and might not be necessary at the solution while adding a regularization term directly to the objective might aid the algorithm but will bias the solution. Addition of regularization terms to the objective function is common in estimation and control studies and should be performed with care.

3 Inertia-Free Setting

The primary practical intention of inertia correction is to guarantee that the direction d_k is of descent. This approach, however, introduces a disconnect between the regularization procedure (IBR) and the filter line-search procedure. In particular, the inertia test (2.10) is based solely on the structural properties of M_k and not on the computed direction d_k. Hence, the regularization procedure IBR might

discard productive descent directions in attempting to enforce a correct inertia. This is particularly problematic in the presence of ill-posed problems.

We recently proposed inertia-free approaches that ensure that the computed step is a descent direction. The first approach requires the tanjential component t_k satisfies the curvature test

$$t_k^T W_k t_k \geq \alpha_{IFR} t_k^T t_k, \tag{3.12}$$

for a small constant $\alpha_{IFR} > 0$. To see the motivation behind this curvature test we first note that a direction $d_k = n_k + t_k$ is of descent if $g_k^T d_k < 0$. We also know that $g_k^T d_k = g_k^T n_k + g_k^T t_k$. If we multiply the first row of (2.8) by t_k we have that,

$$t_k^T W_k t_k + t_k^T J_k^T \lambda_k^+ = -t_k^T g_k + t_k W_k^n$$

adding and substracting $n_k^T g_k$ we obtain,

$$-g_k^T d_k = t_k^T W_k t_k + t_k^T J_k^T \lambda_k^+ t_k^T W_k n_k - g_k^T n_k.$$

From the second row of (2.8) we have that $J_k t_k = 0$ (t_k lies on the nullspace of J_k) and thus $t_k^T J_k^T = 0$. Moreover, from (2.7) we have $n_k = O(\|c_k\|)$ and thus $-g_k^T d_k > 0$ for sufficiently small c_k and $t_k W_k t_k > 0$.

The key observation is that the curvature conditions (3.12) can hold even if M_k does not have correct inertia. Consequently, the curvature condition is less restrictive than the inertia test and can accept productive steps more easily. If the curvature conditions do not hold for the computed step, we can enforce them by regularizing the Hessian matrix $W_k \leftarrow W_k + \delta I$. Specifically, the curvature condition (3.12) can always be satisfied for sufficiently large δ and sufficiently small α_{IFR} satisfying $\alpha_{IFR} \leq \lambda_{min}(Z_k^T W_k Z_k)$ where $\lambda_{min}(A)$ is the minimum eigenvalue of a given matrix A. The reason is that t_k lies on the null space of J_k and, consequently, can always be expressed as $t_k = Z_k u$ for a given nonzero vector u and the curvature condition then implies that $u^T Z_k^T W_k Z_k u \geq \alpha_{IFR} u^T u$. We also know that an appropriate α_{IFR} exists for any δ because $\lambda_{min}(Z_k^T W_k(\delta) Z_k)$ is an increasing function of δ. From (2.9) we can also see that we can enforce nonsingularity of M_k by regularizing the Hessian because the reduced Hessian will be nonsingular for sufficiently large σ which would imply that it does not have zero eigenvalues and thus M_k does not have zero eigenvalues. These observations lead to the following inertia-free regularization (IFRt) procedure:

INERTIA-FREE REGULARIZATION (IFRt)

IFR-1 Given constant $\alpha_{IFR} > 0$, factorize M_k with $\delta = 0$ and compute n_k, t_k from (2.7) and (2.8). If t_k satisfies the curvature condition (3.12) and M_k is nonsingular, set $d_k = t_k + n_k$, and terminate.

IFR-2 If $\delta^{last} = 0$, set $\delta \leftarrow \bar{\delta}^0$, otherwise set $\delta \leftarrow \max\{\delta^{min}, \kappa^- \delta^{last}\}$.

IFR-3 Given constant $\alpha_{IFR} > 0$, factorize M_k with current δ and compute n_k, t_k from (2.7) and (2.8). If t_k satisfies (3.12) and M_k is nonsingular, set $d_k \leftarrow n_k + t_k$, and terminate.

IFR-4 If $\delta^{last} = 0$, set $\delta \leftarrow \hat{\kappa}^+ \delta$, otherwise set $\delta \leftarrow \kappa^+ \delta$ and go to IFR-3.

Iter is also possible to directly operate with the step d_k computed from (2.4). In other words, the step does not need to be decomposed in tangential and normal components. The corresponding curvature test is:

$$d_k^T W_k d_k \geq \alpha_{IFR} d_k^T d_k, \tag{3.13}$$

A similar regularization procedure is used in this case (we denote this approach IFRd). We can see the motivation behind this test by multiplying the first row of (2.4) by d_k^T which gives,

$$-d_k^T g_k = d_k^T W_k d_k + d_k^T J_k^T \lambda_k^+. \tag{3.14}$$

From the second row we have that $J_k d_k = -c_k$ and thus we have that $-d_k^T g_k > 0$ for sufficiently small c_k if $d_k^T W_k d_k > 0$. An advantage of the IFRd procedure is that it can be more flexible than IFRt because it does not require of positive curvature on the null space of J_k to ensure descent. A disadvantage, however, is that the step d_k might be much larger than t_k (i.e., if the normal component n_k is large) and this can introduce numerical difficulties when performing the test (3.13).

4 Case Studies

We present in estimation and control studies to illustrate how ill-posed problem formulations can significantly deteriorate the performance of an inertia-based approach and to demonstrate that inertia-free approaches can help overcome these issues. All the NLPs considered are implemented in the algebraic modeling language `AMPL` [13]. The `AMPL` model and data for these studies can be obtained from `http://zavalab.engr.wisc.edu/data`. All instances are solved using `PIPS-NLP`, an open-source parallel interior-point solver that we use to experiment with different algorithmic and linear algebra strategies [7, 8]. `PIPS-NLP` is available for download at `https://github.com/Argonne-National-Laboratory/PIPS`.

4.1 Estimation of Weakly Observable Systems

We consider a state estimation problem for the nonlinear system:

$$\frac{dx_1}{d\tau} = x_2(\tau) \tag{4.15a}$$
$$\frac{dx_2}{d\tau} = x_1(\tau) + x_2(\tau) + (1 - x_1(\tau)^2)x_3(\tau) + u(\tau) \tag{4.15b}$$
$$\frac{dx_3}{d\tau} = -x_1(\tau) + x_3(\tau). \tag{4.15c}$$

We assume that $x_1(\cdot)$ is the only measured state while $u(\tau)$ are fixed control actions. An analysis of the observability matrix of this system reveals that it becomes unobservable for $x_1 = 1$ and $x_1 = -1$ [4, 31]. The estimator cost function is given by,

$$\min \quad \frac{1}{2}\kappa(x_1(0) - \bar{x}_1^0)^2 + \frac{1}{2}\kappa(x_2(0) - \bar{x}_2^0)^2 + \frac{1}{2}\kappa(x_3(0) - \bar{x}_3^0)^2 + \int_0^T (x_1(\tau) - \bar{x}_1(\tau))^2 d\tau \tag{4.16}$$

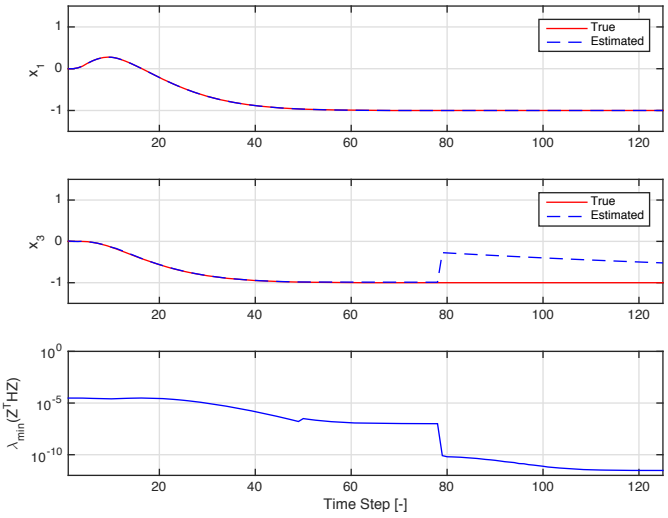

Figure 1: True and estimated trajectories for state x_1 (top graph) and x_3 (middle graph). Jump in minimum eigenvalue of reduced Hessian (bottom graph).

where $\kappa \geq 0$ is the weight of the prior (also known as the arrival cost). The states are required to satisfy the bounds:

$$-5 \leq x_1(\tau) \leq 5 \tag{4.17a}$$

$$-5 \leq x_2(\tau) \leq 5 \tag{4.17b}$$

$$-5 \leq x_3(\tau) \leq 5. \tag{4.17c}$$

The continuous-time problem is transformed into an NLP by using an implicit Euler discretization scheme with N points. The discretization points match the measurement points. We use $\bar{x}^0 = [1\ 1\ 1]$ as the reference prior state.

To illustrate the observability issues that arise in this system we consider the case with $\mu = 0$. In Figure 1 we illustrate the convergence of a moving horizon estimator that solves sequences of estimation problem with an horizon of $N = 5$ [31]. In the top graph we present the convergence to the measured state x_1. In the middle graph we see that the estimator converges to the true state x_3 but at time step 80 we have that x_1 fully settles at -1 and the estimate of x_3 diverges. In the bottom graph we show the minimum eigenvalue of the reduced Hessian matrix. We note the small magnitude of the eigenvalue ($O(10^{-5})$-$O(10^{-10})$), indicating that the system is weakly observable. At time step 80 there is a sudden jump of five orders of magnitude of the eigenvalue from $O(10^{-5})$ to $O(10^{-10})$, which explains the divergence of the x_3.

In Table 1 we summarize the numerical performance of inertia-based (IBR) and inertia-free (IFR) algorithms for increasing values of the estimation horizon N. Here, (Reg) indicates the total number of regularization trials needed. We first consider the case with no prior information ($\kappa = 0.0$). We can see that IBR requires significant amounts of regularization because the system is weakly observable.

This trend is maintained even if a long horizon is used. The IFRd approach can solve all instances but its performance is not competitive. The IFRt approach, on the other hand, reduces the amount of regularization quite dramatically. In this table we can also see that the performance of IBR does not improve as we increase prior regularization. This indicates that prior information and constraints are insufficient to estimate the states uniquely. Moreover, the addition of prior information in fact deteriorates the solver performance. IFRd is again not competitive but IFRt outperforms IBR in all cases. We attribute the difference in performance between IFRd and IFRt to the fact that the normal component of the step dominates that of the tangential component.

Table 1: Number of iterations (Iter) and regularizations (Reg) for IBR and IFR.

		IBR		**IFRd**		**IFRt**	
	N	Iter	Reg	Iter	Reg	Iter	Reg
	25	8	0	8	0	8	0
	50	18	16	23	12	15	8
$\kappa = 0.0$	75	33	17	52	28	30	6
	100	34	22	49	21	34	5
	125	46	22	74	46	47	9
	150	52	21	83	62	51	8
$\kappa = 0.1$	150	54	21	Fail	Fail	54	8
$\kappa = 1.0$	150	56	21	71	71	55	15
$\kappa = 10$	150	62	29	97	111	59	13

4.2 EVM Estimation

We consider the problem of inferring the number of occupants $n_{oc}(\tau)$ and total inlet air flow $q_{in}(\tau)$ in a space of volume V using sensor signals of carbon dioxide $C(\tau)$. In Figure 2 we present real signals for carbon dioxide, ambient flows, and estimated occupancy rates obtained with moving horizon estimation [32].

The building-wide mass balance of the system is given by

$$\frac{1}{\rho(\tau)}\frac{dm}{dt}(\tau) = q_{in}(\tau) - q_{out}(\tau) + q_{inf}(\tau) - q_{exf}(\tau) \tag{4.18a}$$

$$V\frac{dC}{dt}(\tau) = q_{in}(\tau)\cdot C_{in}(\tau) - q_{out}(\tau)\cdot C(\tau) + q_{inf}(\tau)\cdot C_{in}(\tau) - q_{exf}\cdot C(\tau) + G\cdot n_{oc}(\tau). \tag{4.18b}$$

The total air mass is given by $m(\tau)$. The inlet concentration $C_{in}(\tau)$ is the atmospheric concentration and is assumed to be constant. The outlet, infiltration, and exfiltration flows are denoted by $q_{out}(\tau), q_{inf}(\tau)$ and $q_{exf}(\tau)$, respectively. G is the average CO_2 generation rate per occupant, and $\rho(\tau)$ is the air density.

We assume that an internal feedback controller maintains the system pressure constant, which implies that the mass $m(\tau)$ is constant and $0 = q_{in}(\tau) - q_{out}(\tau) + q_{inf}(\tau) - q_{exf}(\tau)$. Using this relationship, we have that

$$V\frac{dC}{dt}(\tau) = (q_{in}(\tau) + q_{inf}(\tau))\cdot(C_{in}(\tau) - C(\tau)) + G\cdot n_{oc}(\tau). \tag{4.19}$$

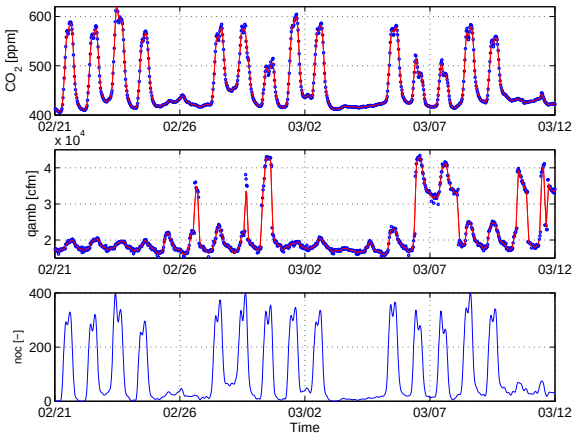

Figure 2: Carbon dioxide and ambient flow signals (top two panels). Estimated occupancy (bottom panel).

Note that the exfiltration and outlet flows do not appear in the balance equation under this assumption. Buildings are typically operated under positive pressure (internal building pressure higher than atmospheric) [30, 21], effectively minimizing infiltration. Consequently, one can assume that $q_{inf}(\cdot) \approx 0$. Note that this assumption does not hold in the case of natural ventilation [24]. With this, we have that

$$V\frac{dC}{dt}(\tau) = q_{in}(\tau) \cdot (C_{in}(\tau) - C(\tau)) + G \cdot n_{oc}(\tau). \tag{4.20}$$

We assume that sensor measurements for $C(\tau)$ are available while the unknown signals are $n_{oc}(\tau)$ and $q_{in}(\tau)$. This gives an errors-in-variables-measured (EVM) formulation in which the system inputs and states are estimated. Note also that $n_{oc}(\tau)$ can be interpreted as process noise signals. The sensor signals are assumed to follow Gaussian distributions $C(\tau) \sim \mathcal{N}(\bar{C}(\tau), \sigma_C^2)$ with known and constant variances $\sigma_{q_{in}}^2, \sigma_C^2$. The prior initial condition at time $t-T$, where T is the estimation horizon is also assumed to be Gaussian $C(t-T) \sim \mathcal{N}(\bar{C}(t-T), \sigma_C^2)$ as this is also measured. The estimator uses the log-likelihood function:

$$\min \quad \int_{t-T}^{t} \left(\frac{1}{2\sigma_C^2} (C(\tau) - \bar{C}(\tau))^2 \right) d\tau. \tag{4.21}$$

We constrain the estimator to the physical CO_2 dynamics (4.20) defined over $\tau \in [t-T, t]$. We use the following constraints

$$L_C(\tau) \leq C(\tau) \leq U_C(\tau) \tag{4.22a}$$

$$L_{n_{oc}}(\tau) \leq n_{oc}(\tau) \leq U_{n_{oc}}(\tau) \tag{4.22b}$$

$$L_{q_{in}}(\tau) \leq q_{in}(\tau) \leq U_{q_{in}}(\tau), \tag{4.22c}$$

for $\tau \in [t-T, t]$. Here, $L_i(\tau), U_i(\tau)$ are lower- and upper-bound trajectories with $i \in \{C, n_{oc}, q_{in}\}$.

Regularization provides a mechanism to incorporate information about the estimated quantities in order to mitigate ill-posedness. As we have discussed, in an estimation context this can be done

by using using prior terms. From the structure of the augmented matrix (2.11) we have also seen that curvature can be introduced implicitly by using bound constraints. Bound constraints also help filter out regions of high probability with no physical meaning (e.g., negative concentrations, unrealistic occupancy levels) [15]. In the case of estimation of infinite-dimensional signals such as $n_{oc}(\tau)$ and $q_{amb}(\tau)$, however, constraints and prior terms might be insufficient to eliminate ill-posedness, as we now illustrate.

In Table 5 we present the numerical performance of IFR and IBR for occupancy estimation problems with increasing horizons N and we changing parameters α_{IFR} for the inertia-free curvature tests. We can see that, for a short horizon of $N = 50$ the measured data is sufficient to estimate the $n_{oc}(\cdot)$ and $q_{amb}(\cdot)$ signals, indicated by the lack of regularization needed by IBR. Note that IFRd and IFRt correctly detect this. As we increase the horizon the measured data becomes insufficient to estimate the signals. This is in part given by the increase in the number of degrees of freedom. We can see that IFRd and IFRt behave exactly the same in all instances, indicating that the normal component of the step is small. We also see that IFRd and IFRd dramatically outperform IBR both in the number of iterations an din the number of regularizations needed. This illustrates that the ability to reduce the amount of regularization can accelerate the search. When we increase α_{IFR} from 1×10^{-11} to 1×10^{-10} for IFRd and IFRt we can see that the number of iterations is reduced but the number of regularizations is increased. This is because increasing α_{IFR} has the effect of requesting for larger positive curvature for the step. On the other hand, when we decrease α_{IFR} to 1×10^{-12} the curvature test is more forgiving, and requests less regularization but at the expense of increasing the number of iterations. The ability to tune α_{IFR} provides flexibility to the inertia-free strategies while the inertia-based approach does not have such flexibility. We can also see that the objective values obtained with both IFR and IBR are close, indicating that IFR is as reliable as IBR.

Table 2: Number of iterations and additional regularizations for IBR and IFR.

		IBR			**IFRd**			**IFRt**		
	N	Obj	Iter	Reg	Obj	Iter	Reg	Obj	Iter	Reg
$\alpha_{IFR} = 1 \times 10^{-11}$	50	1.54×10^{1}	27	0	1.54×10^{1}	27	0	1.54×10^{1}	27	0
	100	3.16×10^{1}	51	30	3.17×10^{1}	29	1	3.17×10^{1}	29	1
	500	6.73×10^{1}	66	23	6.73×10^{1}	43	2	6.74×10^{1}	43	2
	1000	8.48×10^{1}	69	29	8.49×10^{1}	44	2	8.49×10^{1}	44	2
	1250	9.93×10^{1}	100	73	9.94×10^{1}	81	25	9.94×10^{1}	81	25
$\alpha_{IFR} = 1 \times 10^{-10}$	1250	9.93×10^{1}	100	73	9.93×10^{1}	79	41	9.93×10^{1}	79	41
$\alpha_{IFR} = 1 \times 10^{-12}$	1250	9.93×10^{1}	100	73	9.94×10^{1}	82	24	9.94×10^{1}	82	24

4.3 Stochastic Optimal Control

We now consider a large-scale optimal control study to demonstrate that the inertia-free approach can handle large and highly nonlinear applications. Moreover, we use this study to illustrate how economic objectives give rise to ill-posed formulations that are difficult to handle numerically.

4.3.1 Problem Formulation

Transport Equations. We use the following form of the continuity, momentum, and energy conservation equations for a pipeline segment $\ell \in \mathcal{L}$:

$$\frac{\partial p_\ell(x,\tau,\omega)}{\partial \tau} + \frac{ZRT_\ell(x,\tau,\omega)}{A_\ell}\frac{\partial f_\ell(x,\tau,\omega)}{\partial x} = 0 \tag{4.23a}$$

$$\frac{1}{A_\ell}\frac{\partial f_\ell(x,\tau,\omega)}{\partial \tau} + \frac{\partial p_\ell(x,\tau,\omega)}{\partial x} + \frac{8\lambda_\ell}{\pi^2 D_\ell^5}\frac{f_\ell(x,\tau,\omega)|f_\ell(x,\tau,\omega)|}{\rho_\ell(x,\tau,\omega)} = 0 \tag{4.23b}$$

$$\begin{aligned}\rho(x,\tau,\omega)\, c_p \left(\frac{\partial T_\ell(x,\tau,\omega)}{\partial \tau} + \nu(x,\tau,\omega)\frac{\partial T_\ell(x,\tau,\omega)}{\partial x}\right) \\ - \left(\frac{\partial p_\ell(x,\tau,\omega)}{\partial \tau} + \nu(x,\tau,\omega)\frac{\partial p_\ell(x,\tau,\omega)}{\partial x}\right) \\ + \frac{\pi D_\ell U_\ell}{A_\ell}\left(T_\ell(x,\tau,\omega) - T_{amb}(x,\tau,\omega)\right) = 0\end{aligned} \tag{4.23c}$$

Here, $\tau \in \mathcal{T} := [0,T]$ is the time dimension with final time T, and $x \in \mathcal{X}_\ell := [0, L_\ell]$ is the axial dimension with length L_ℓ and $\omega \in \Omega$ is the scenario set. The link diameters are denoted as D_ℓ and the friction coefficients as λ_ℓ. The states of the link are the gas density $\rho_\ell(x,\tau,\omega)$, the speed $\nu_\ell(x,\tau,\omega)$, the pressure $p_\ell(x,\tau,\omega)$, and the temperature $T_\ell(x,\tau,\omega)$. The ambient temperature is denoted by $T_{amb}(x,\tau,\omega)$ and assumed to be given. The gas pressure and density are related as

$$\frac{p_\ell(x,\tau,\omega)}{\rho_\ell(x,\tau,\omega)} = ZRT(x,\tau,\omega), \tag{4.24}$$

where R is the universal gas constant. The pipeline links are connected through a network comprising a set $\mathcal{N}$ of nodes, a set $\mathcal{S}$ of supply flows, and a set $\mathcal{D}$ of demand flows. For each node $n \in \mathcal{N}$ we define the set of inlet and outlet links, $\mathcal{L}_n^{snd} := \{\ell \,|\, snd(\ell) = n\}, \mathcal{L}_n^{rec} := \{\ell \,|\, rec(\ell) = n\}$. Here, $snd(\ell) \in \mathcal{N}$ is the start node of link ℓ and $rec(\ell) \in \mathcal{N}$ is the end node. We assume that the direction of the flow is given. We define $dem(j) \in \mathcal{N}$ as the node at which the demand flow $d_j(\tau,\omega)$ is located and $sup(i) \in \mathcal{N}$ as the node at which the supply flow $s_i(\tau,\omega)$ is located. Accordingly, we define the sets $\mathcal{S}_n := \{j \in \mathcal{S} \,|\, sup(j) = n\}$ and $\mathcal{D}_n := \{j \in \mathcal{D} \,|\, dem(j) = n\}$ for each node $n \in \mathcal{N}$. For each node $n \in \mathcal{N}$ we also define pressures $\theta_n(\tau,\omega)$ and temperatures $\phi_n(\tau,\omega)$.

For modeling convenience we lift the network system by introducing *dummy* inlet flow and enthalpies for each link $f_\ell^{in}(\tau,\omega), h_\ell^{in}(\tau,\omega)$, respectively; and outlets $f_\ell^{out}(\tau,\omega), h_\ell^{out}(\tau,\omega)$. Using these definitions we can express flow and enthalpy balances at the nodes as

$$\sum_{\ell\in\mathcal{L}_n^{rec}} f_\ell^{out}(\tau,\omega) - \sum_{\ell\in\mathcal{L}_n^{snd}} f_\ell^{in}(\tau,\omega) + \sum_{i\in\mathcal{S}_n} s_i(\tau,\omega) - \sum_{j\in\mathcal{D}_n} d_j(\tau,\omega) = 0,\ n \in \mathcal{N} \tag{4.25a}$$

$$\sum_{\ell\in\mathcal{L}_n^{rec}} h_\ell^{out}(\tau,\omega) - \sum_{\ell\in\mathcal{L}_n^{snd}} h_\ell^{in}(\tau,\omega) + \sum_{i\in\mathcal{S}_n} h_i^{s}(\tau,\omega) - \sum_{j\in\mathcal{D}_n} h_j^{d}(\tau,\omega) = 0,\ n \in \mathcal{N}. \tag{4.25b}$$

Where,

$$h_\ell^{in}(\tau,\omega) = c_p f_\ell^{in}(\tau,\omega)\phi_{snd(\ell)}(\tau,\omega),\ \ell \in \mathcal{L} \tag{4.26a}$$
$$h_\ell^{out}(\tau,\omega) = c_p f_\ell^{out}(\tau,\omega)T_\ell(L_\ell,\tau,\omega),\ \ell \in \mathcal{L} \tag{4.26b}$$
$$h_i^s(\tau,\omega) = c_p s_i(\tau,\omega)T_i^s,\ i \in \mathcal{S} \tag{4.26c}$$
$$h_j^d(\tau,\omega) = c_p d_j(\tau,\omega)T_{dem(j)}(\tau,\omega),\ j \in \mathcal{D}, \tag{4.26d}$$

and T_i^s is the supply gas temperature (assumed to be given). We split the set of links $\mathcal{L}$ into subsets of passive $\mathcal{L}_p$ links and active links $\mathcal{L}_a$. For the active links we define the *boost pressures* $\Delta\theta_\ell(\tau,\omega)$ which are the additional (non-negative) pressures introduced by the compressor located at the inlet (sending) node of the link. For the passive links there is no compression. The boundary conditions for the link pressures are given by

$$p_\ell(L_\ell,\tau,\omega) = \theta_{rec(\ell)}(\tau,\omega),\ \ell \in \mathcal{L} \tag{4.27a}$$
$$p_\ell(0,\tau,\omega) = \theta_\ell^{dis}(\tau,\omega),\ \ell \in \mathcal{L}. \tag{4.27b}$$

The boundary conditions for the link flows and temperatures are,

$$f_\ell(0,\tau,\omega) = f_\ell^{in}(\tau,\omega),\ \ell \in \mathcal{L} \tag{4.28a}$$
$$f_\ell(L_\ell,\tau,\omega) = f_\ell^{out}(\tau,\omega),\ \ell \in \mathcal{L} \tag{4.28b}$$
$$T_\ell(0,\tau,\omega) = \phi_\ell^{dis}(\tau,\omega),\ \ell \in \mathcal{L}. \tag{4.28c}$$

The discharge pressures and temperatures of the compressors are given by,

$$\theta_\ell^{dis}(\tau,\omega) = \theta_{snd(\ell)}(\tau,\omega) + \Delta\theta_\ell(\tau,\omega),\ \ell \in \mathcal{L}_a \tag{4.29a}$$
$$\phi_\ell^{dis}(\tau,\omega) = \phi_{snd(\ell)}(\tau,\omega)\left(\frac{\theta_\ell^{dis}(\tau,\omega)}{\theta_{snd(\ell)}(\tau,\omega)}\right)^\beta,\ \ell \in \mathcal{L}_a. \tag{4.29b}$$

For the passive links we simply have that,

$$\theta_\ell^{dis}(\tau,\omega) = \theta_{snd(\ell)}(\tau,\omega),\ \ell \in \mathcal{L}_p \tag{4.30a}$$
$$\phi_\ell^{dis}(\tau,\omega) = \phi_{snd(\ell)}(\tau,\omega),\ \ell \in \mathcal{L}_p. \tag{4.30b}$$

The total compression power consumed in the active links is given by

$$P_\ell(\tau,\omega) = c_p\, f_\ell^{in}(\tau,\omega)\, \phi_{snd(\ell)}(\tau,\omega)\left(\left(\frac{\theta_\ell^{dis}(\tau,\omega)}{\theta_{rec(\ell)}(\tau,\omega)}\right)^\beta - 1\right),\ \ell \in \mathcal{L}_a. \tag{4.31}$$

From (4.29) we note that this last expression can be written as,

$$P_\ell(\tau,\omega) = c_p\, f_\ell^{in}(\tau,\omega)(\phi_\ell^{dis}(\tau,\omega) - \phi_{snd(\ell)}(\tau,\omega)),\ \ell \in \mathcal{L}_a. \tag{4.32}$$

Constraints. We impose the following constraints on available compressor power, suction and discharge pressures, discharge temperatures, and demand delivery pressures:

$$P_\ell^L \leq P_\ell(\tau,\omega) \leq P_\ell^U,\ \ell \in \mathcal{L}_a \tag{4.33a}$$

$$\theta_\ell^{suc,L} \leq \theta_{snd(\ell)}(\tau,\omega) \leq \theta_\ell^{suc,U},\ \ell \in \mathcal{L}_a \tag{4.33b}$$

$$\theta_\ell^{dis,L} \leq \theta_\ell^{dis}(\tau,\omega) \leq \theta_\ell^{dis,U},\ \ell \in \mathcal{L}_a \tag{4.33c}$$

$$\phi_\ell^{dis,L} \leq \phi_\ell^{dis}(\tau,\omega) \leq \phi_\ell^{dis,U},\ \ell \in \mathcal{L}_a \tag{4.33d}$$

$$\theta_j^{dem,L} \leq \theta_{dem(j)}(\tau,\omega) \leq \theta_j^{dem,U},\ j \in \mathcal{D}. \tag{4.33e}$$

We also use the implicit physical bounds $f_\ell^{in}(\tau,\omega), f_\ell^{out}(\tau,\omega), P_\ell(\tau,\omega) \geq 0$. The gas demands are bounded by the given targets $d_j^{target}(\tau,\omega)$,

$$0 \leq d_j(\tau,\omega) \leq d_j^{target}(\tau,\omega),\ j \in \mathcal{D}. \tag{4.34}$$

We highlight that the demand targets are the source of uncertainty in the problem. We enforce the following periodicity constraint to ensure that gas is replenished at the end of the horizon,

$$\int_0^{L_\ell} f_\ell(x,T,\omega)dx \geq \int_0^{L_\ell} f_\ell(x,0,\omega)dx, \quad \ell \in \mathcal{L}_a. \tag{4.35}$$

Without this constraint, the system will tend to deplete linepack in order to minimize compressor power and this can make operation in the following day infeasible. The periodicity constraint thus provides a mechanism to deal with the finite horizon of the optimal control problem. We assume that the system is at steady-state at the initial time $\tau = 0$. The steady-state satisfies the transport equations

$$\frac{ZRT_\ell(x,\tau,\omega)}{A_\ell}\frac{\partial f_\ell(x,\tau,\omega)}{\partial x} = 0 \tag{4.36a}$$

$$\frac{\partial p_\ell(x,\tau,\omega)}{\partial x} + \frac{8\lambda_\ell}{\pi^2 D_\ell^5}\frac{f_\ell(x,\tau,\omega)|f_\ell(x,\tau,\omega)|}{\rho_\ell(x,\tau,\omega)} = 0 \tag{4.36b}$$

$$\begin{aligned}\rho(x,\tau,\omega)\, c_p\, \nu(x,\tau,\omega)\frac{\partial T_\ell(x,\tau,\omega)}{\partial x} - \nu(x,\tau,\omega)\frac{\partial p_\ell(x,\tau,\omega)}{\partial x} \\ + \frac{\pi D_\ell U_\ell}{A_\ell}\left(T_\ell(x,\tau,\omega) - T_{amb}(x,\tau,\omega)\right) = 0\end{aligned} \tag{4.36c}$$

Using an analysis similar to that presented in [33] it is possible to prove that the only *degrees of freedom* of the problem are the controls $\Delta\theta_\ell(\tau,\omega),\ \ell \in \mathcal{L}_a, \tau \in \mathcal{T}$ and demands $d_j(\tau,\omega), j \in \mathcal{D}, \tau \in \mathcal{T}$. Moreover, we can prove that fixing $\Delta\theta_\ell(0,\omega)$ and $d_j(0,\omega)$ fully defines the initial states for each scenario $\omega \in \Omega$.

Nonanticipativity. Iter is possible to consider different decision-making settings depending on how information is presented to the operator. Here we consider a preparation period $\mathcal{T}_{prep} := [0, T_{prep}]$ with $T_{prep} < T$ to allocate linepack and we seek to find out how much linepack to allocate to withstand uncertain variations of demand past the preparation period. This setting can be modeled by imposing the non-anticipativity constraints:

$$\Delta\theta_\ell(\tau,\omega) = \mathbb{E}\left[\Delta\theta_\ell(\tau,\omega)\right],\ \tau \in \mathcal{T}_{prep} \tag{4.37}$$

$$d_j(\tau,\omega) = \mathbb{E}\left[d_j(\tau,\omega)\right],\ \tau \in \mathcal{T}_{prep}. \tag{4.38}$$

Objective Functions. We use the expected cost $\mathbb{E}[\varphi(\omega)]$ as summarizing statistic for random objective function. The per-scenario cost $\varphi(\omega)$ is a combination of compression cost and gas demand delivery,

$$\varphi(\omega) := \int_0^T \left(\sum_{\ell \in \mathcal{L}_a} \alpha_\ell^P P_\ell(\tau, \omega) - \sum_{j \in \mathcal{D}} \alpha_j^d d_j(\tau, \omega) \right) d\tau \tag{4.39}$$

We note that the first term in the cost function is the compressor power with cost α_ℓ^P and the second term is the total value of served demand. Specifically, the second term seeks to maximize the served demand (this explains the negative sign) and the parameter α_j^d can be interpreted as the price of the served load. When the value α_j^d is high relative to compression power (as is often the case), the system will tend to push the demands $d_j(\tau, \omega)$ to the demand targets $d_j^{target}(\tau, \omega)$ defined in the constraints (4.34). We highlight that this cost function is linear and, as we will show in Section 4.4, introduces ill-conditioning and drastically slows down convergence of the inertia-based approach IBR. To ameliorate this effect a quadratic demand tracking term can also be used [33]:

$$\varphi(\omega) := \int_0^T \left(\sum_{\ell \in \mathcal{L}_a} \alpha_\ell^P P_\ell(\tau, \omega) + \sum_{j \in \mathcal{D}} \alpha_j^d (d_j(\tau, \omega) - d_j^{target}(\tau, \omega))^2 \right) d\tau. \tag{4.40}$$

An advantage of this formulation is that it adds positive curvature to the cost function. A disadvantage of this formulation is that the parameter α_j^d does not have a direct economic interpretation and it might thus be difficult to tune. This raises an important point that is often made in the economic MPC literature: the direct use of economic objectives is attractive from a performance standpoint but tends to make problems harder to solve while tracking terms are easier to solve but deteriorate performance if not properly tuned [34]. With the inertia-free approaches we seek to enable more efficient solutions of problems with economic objectives.

4.3.2 Problem Decomposition

After discretization, the stochastic optimal control problem can be cast as the structured NLP:

$$\begin{aligned}
\min \quad & f_0(x_0) + \sum_{\omega \in \Omega} f_\omega(x_\omega, x_0) && && \text{(4.41a)} \\
\text{s.t.} \quad & c_0(x_0) = 0 && (\lambda_0) && \text{(4.41b)} \\
& c_\omega(x_\omega, x_0) = 0, \ \omega \in \Omega && (\lambda_\omega) && \text{(4.41c)} \\
& x_0 \geq 0 && (\nu_0) && \text{(4.41d)} \\
& x_\omega \geq 0, \ \omega \in \Omega && (\nu_\omega) && \text{(4.41e)}
\end{aligned}$$

We denote this NLP as a structured problem with *primal coupling*. Here, $x_\omega \in \Re^{n_\omega}$ are the per-scenario variables and the interface or coupling variables are $x_0 \in \Re^{n_0}$. The dual variables are $\lambda_0 \in \Re^{m_0}$ and $\lambda_\omega \in \Re^{m_\omega}$. The augmented system of (4.41) can be permuted into the block-bordered-diagonal (BBD)

form,

$$\left[\begin{array}{c|cccc} K_0 & B_1^T & B_2^T & \dots & B_{|\Omega|}^T \\ \hline B_1 & K_1 & & & \\ B_2 & & K_2 & & \\ \vdots & & & \ddots & \\ B_{|\Omega|} & & & & K_{|\Omega|} \end{array}\right] \left[\begin{array}{c} \Delta w_0 \\ \Delta w_1 \\ \Delta w_2 \\ \vdots \\ \Delta w_{|\Omega|} \end{array}\right] = - \left[\begin{array}{c} r_0 \\ r_1 \\ r_2 \\ \vdots \\ r_{|\Omega|} \end{array}\right], \tag{4.42}$$

where $\Delta w_0 = (\Delta x_0, \lambda_0^+)$, $\Delta w_\omega = (\Delta x_\omega, \lambda_\omega^+)$,

$$K_0 = \left[\begin{array}{cc} W_0(\delta) & J_0^T \\ J_0 & \end{array}\right], \tag{4.43a}$$

$$K_\omega = \left[\begin{array}{cc} W_\omega(\delta) & J_\omega^T \\ J_\omega & \end{array}\right], \tag{4.43b}$$

$$B_\omega^T = \left[Q_\omega^T \; T_\omega^T\right], \tag{4.43c}$$

$J_0 = \nabla_{x_0} c_0(x_0)$, $J_\omega = \nabla_{x_\omega} c_\omega(x_0, x_\omega)$, $T_\omega = \nabla_{x_0} c_\omega(x_0, x_\omega)$, $W_0(\delta) = \nabla_{x_0,x_0}\mathcal{L} + X_0^{-1}V_0 + \delta\mathbb{I}_{n_0}$, $W_\omega(\delta) = \nabla_{x_\omega,x_\omega}\mathcal{L} + X_\omega^{-1}V_\omega + \delta\mathbb{I}_{n_\omega}$, and $Q_\omega = \nabla_{x_0,x_\omega}\mathcal{L}$. Here,

$$\mathcal{L} := f_0(x_0) + \lambda_0^T c_0(x_0) + \sum_{\omega\in\Omega} \left(f_\omega(x_\omega, x_0) + \lambda_\omega^T c_\omega(x_\omega, x_0)\right). \tag{4.44}$$

Iter is well-known that the BBD system (4.42) can be solved in parallel by using a Schur complement decomposition approach [19, 18]. This requires the solution of the following systems:

$$\left(K_0 - \sum_{\omega\in\Omega} B_\omega K_\omega^{-1} B_\omega^T\right) \Delta w_0 = -r_0 + \sum_{\omega\in\Omega} K_\omega^{-1} B_\omega r_\omega \tag{4.45a}$$

$$K_\omega \Delta w_\omega = -r_\omega - B_\omega^T \Delta w_0, \; \omega \in \Omega. \tag{4.45b}$$

Here, $S{:=}K_0 - \sum_{\omega\in\Omega} B_\omega K_\omega^{-1} B_\omega^T$ is the *Schur complement matrix*. This is a symmetric and indefinite matrix of dimension $n_0{+}m_0$. In the inertia-based approach IBR, we check the inertia of the BBD system by using Haynsworth's formula:

$$\text{Inertia}(M_k(\delta)) = \sum_{\omega\in\Omega} \text{Inertia}(K_\omega) + \text{Inertia}\left(K_0 - \sum_{\omega\in\Omega} B_\omega K_\omega^{-1} B_\omega^T\right), \tag{4.46}$$

and we recall that

$$n = n_0 + \sum_{\omega\in\Omega} n_\omega, \quad m = m_0 + \sum_{\omega\in\Omega} m_\omega. \tag{4.47}$$

Consequently, if we have that $\text{Inertia}(K_\omega){=}\{n_\omega, m_\omega, 0\}$ for all $\omega \in \Omega$ then the inertia of $M_k(\delta)$ is correct if and only if $\text{Inertia}(S) = \{n_0, m_0, 0\}$. One can obtain the inertia of the blocks K_ω and of S by using LBL^T factorizations.

4.4 Numerical Results

We consider a pipeline with $|\mathcal{N}| = 13$ nodes, $|\mathcal{L}| = 12$ links, $|\mathcal{L}_a| = 10$ compressors, $|\mathcal{S}| = 1$ supply flow located at the first node, and $|\mathcal{D}| = 1$ demand flow located at the last node. The system spans 1,600 km with 36 inch (914 mm) pipes. The distance between compressors is 100 km. We consider a planning horizon T of 24 hours discretized in $N_t = 48$ time intervals $\Delta\tau$ of 30 min (1800 sec). Each link is discretized in space by using $N_x = 3$ points. We consider a demand process that simulates strong uncertainty in the start and end of a process consuming gas (e.g., a power plant or refinery). The scenarios are illustrated in Figure 3.

We solve the instances using the Schur decomposition implementation of `PIPS-NLP` and we perform LBL^T factorizations of the scenario blocks using the sparse linear solver `MA57`. We use the nested dissection strategy implemented in `METIS` to perform reordering [12, 20, 36, 37]. All parallel experiments were performed on the Fusion computing cluster at Argonne National Laboratory. Fusion contains 320 computing nodes, and each node has two quad-core Nehalem 2.6 GHz Pentium Xeon CPUs.

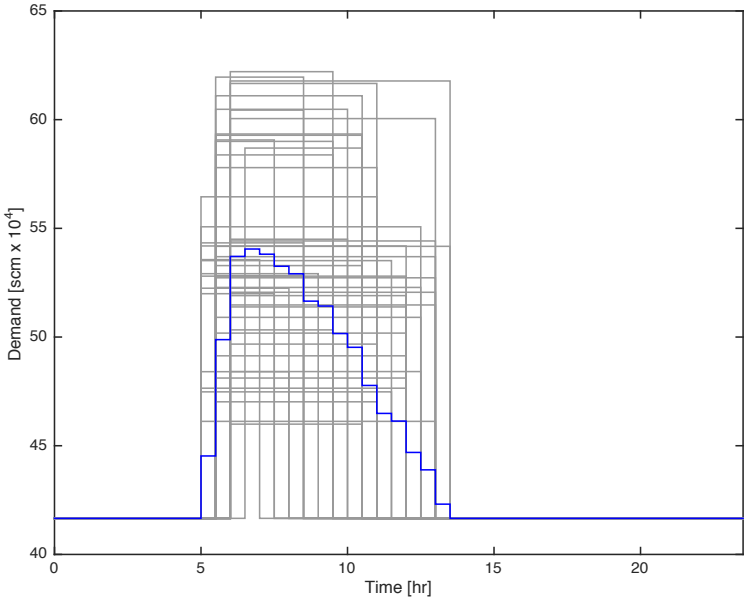

Figure 3: Demand scenarios (gray) and mean process (blue).

We consider a stochastic nonisothermal control formulation with up to 128 scenarios. The resulting NLP with linear economic objective (4.39) (that we call `GAS_Nonisoth_Lin`) contains 1,799,051 variables, 1,766,400 equality constraints and 37,120 inequality constraints. The NLP with quadratic tracking costs (4.40) (that we call `GAS_Nonisoth_Quad`) has 1,798,795 variables, 1,766,400 equality constraints and 30,848 inequality constraints. To demonstrate the strong effect that nonsiothermal behavior has on the solver performance we also consider an stochastic isothermal control formulation with a linear economic objective `GAS_Isoth_Lin`. This NLP has 843,659 variables, 811,008 equality constraints, 34,048 inequality constraints. We set the inertia-free curvature parameter $\alpha_{IFR} =$

1×10^{-11} and we use a default convergence tolerate of 1×10^{-6}.

We first illustrate that the use of quadratic and linear cost functions has significant impacts on economic performance. To do so, we use a one scenario instance. In Table 3 we compare the total compressor power and the load error obtained with the different formulations. We can see that compressor power is underestimated by nearly 30% with the quadratic tracking formulation and that there is a significant tracking error in the load. The difference in the compressor power time profile is illustrated in Figure 4 while in Figure 5 we illustrate the difference in the resulting axial temperature profile. While it is possible to tune the quadratic tracking term to match the performance of the linear economic formulation, such tuning is cumbersome.

Table 3: Pipeline performance with quadratic and linear economic objectives

	Total Power [kWh]	Load Error [scm x 10^4]
Quad	43,295	71.84
Lin	61,767	0

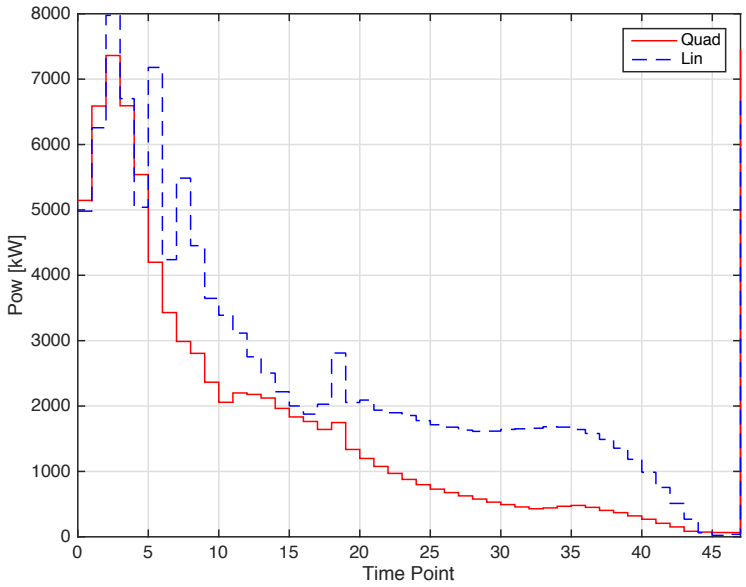

Figure 4: Total compressor power profiles for quadratic and linear cost functions.

We now compare the numerical performance of `PIPS-NLP` in the solution of the different problem instances with up to 128 scenarios. The results are presented in Table 4. Here, (Fact) denotes the total number of factorizations and equals (Reg)+(Iter) while (#MPI) indicates the number of parallel message-passing-interface processes. By comparing `GAS_Isoth_Lin` and `GAS_Nonisoth_Lin` we see that the isothermal formulation requires the same number of iterations (denoted as Iter) and factorizations (denoted as Fact). This indicates that no regularization was needed. We also note that the inertia-free approaches correctly identify this to be the case as well. In the nonisothermal formulations we see a dramatic increase in the number of iterations and solution time. This is particularly

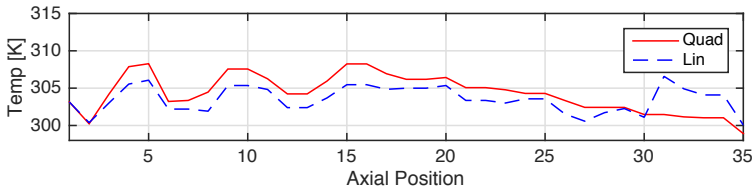

Figure 5: Axial temperature profile for quadratic and linear cost functions.

Table 4: Objectives, iterations (Iter), and total number of factorizations (Fact) for IBR and IFR strategies.

		IBR				IFRd			IFRt		
Problem	#MPI	Obj	Iter	Fact	Time(s)	Iter	Fact	Time(s)	Iter	Fact	Time(s)
GAS_Isoth_Lin	8	-1.03	77	77	146	77	77	146	77	77	131
GAS_Isoth_Lin	16	-1.03	77	77	74	77	77	74	76	76	66
GAS_Isoth_Lin	32	-1.03	77	77	39	77	77	39	77	77	35
GAS_Isoth_Lin	64	-1.03	77	77	22	77	77	22	75	75	19
GAS_Isoth_Lin	128	-1.03	77	77	14	77	77	14	77	77	12
GAS_Nonisoth_Lin	8	-0.97	1196	2352	23256	436	997	7177	753	1803	11184
GAS_Nonisoth_Lin	16	-0.97	1321	2602	13938	436	997	3633	722	1720	5353
GAS_Nonisoth_Lin	32	-0.97	2500	4974	13624	436	997	1829	718	1709	2785
GAS_Nonisoth_Lin	64	-0.97	1097	2154	2948	436	997	957	735	1753	1444
GAS_Nonisoth_Lin	128	-0.97	1138	2236	1642	436	997	522	757	1813	806
GAS_Nonisoth_Quad	8	4.54E4	425	442	5432	246	321	2988	221	268	2416
GAS_Nonisoth_Quad	16	4.54E4	425	442	2729	246	321	1614	228	278	1293
GAS_Nonisoth_Quad	32	4.54E4	425	442	1405	246	321	756	228	278	643
GAS_Nonisoth_Quad	64	4.54E4	425	442	744	246	321	400	218	262	322
GAS_Nonisoth_Quad	128	4.54E4	425	442	418	246	321	220	226	276	190

evident in the performance IBR which requires up to 6.46 hr of solution time on 8 processors and 0.45 hr on 128 processors. We note that the number of factorizations is nearly twice the number of iterations, indicating that one additional regularization step is needed in each iteration. The IFR approaches significantly improve the number of iterations and solution time. In particular, the solution time on 128 processors is 0.14 hr for IFRd and 0.22 hr for IFRt. We note that both IFRt and IFRd need to regularize the Hessian at least once per iteration. The number of iterations, however, is reduced by a factor of nearly two in all cases.

By comparing `GAS_Nonisoth_Lin` and `GAS_Nonisoth_Quad` we see that adding curvature to the cost function by using tracking terms has a dramatic effect on solver performance. The number of iterations are reduced by a factor of five compared to the formulation with linear economic objectives. Moreover, the need for regularization is almost eliminated (this is particularly evident in the IBR case). Interestingly, even in the formulation with quadratic objectives, IFR approaches outper-

form IBR. This again indicates that better quality steps are taken as a result of reduced amounts of regularization. We also highlight that performance of the solver is quite robust and can handle this challenging formulation with strong nonlinearities.

We performed an additional experiment with a reduced tolerance of 1×10^{-5} (we refer to this instance as `GAS_Nonisoth_Lin*`). In Table 5 we can see that no regularization is needed in this instance. This indicates that most of the regularization is observed close to the solution. This is a clear indication that the problem is ill-posed and exhibits solution multiplicity (different combinations of variables achieve the same objective function), explaining the large number of iterations required to close the tolerance.

Table 5: Performance of IBR and IFR strategies under relaxed tolerance.

		IBR				IFRd			IFRt		
Problem	#MPI	Obj	Iter	Fact	Time(s)	Iter	Fact	Time(s)	Iter	Fact	Time(s)
GAS_Nonisoth_Lin	8	-0.97	1196	2352	23256	436	997	7177	753	1803	11184
GAS_Nonisoth_Lin	16	-0.97	1321	2602	13938	436	997	3633	722	1720	5353
GAS_Nonisoth_Lin	32	-0.97	2500	4974	13624	436	997	1829	718	1709	2785
GAS_Nonisoth_Lin	64	-0.97	1097	2154	2948	436	997	957	735	1753	1444
GAS_Nonisoth_Lin	128	-0.97	1138	2236	1642	436	997	522	757	1813	806
GAS_Nonisoth_Lin*	8	-1.08	69	69	491	69	69	485	69	71	451
GAS_Nonisoth_Lin*	16	-1.08	69	69	246	69	69	246	69	69	222
GAS_Nonisoth_Lin*	32	-1.08	69	69	124	69	69	123	69	69	112
GAS_Nonisoth_Lin*	64	-1.08	69	69	66	69	69	66	69	69	59
GAS_Nonisoth_Lin*	128	-1.08	69	69	38	69	69	36	70	72	35

5 Conclusions

We analyzed the performance of an inertia-free, filter line-search approach for solving ill-posed estimation and control problems. We demonstrate that an inertia-free approach can provide flexibility to handle these problems more efficiently, compared to a standard inertia-based approach. As part of future work we will derive hybrid strategies that can exploit both inertia information and inertia-free curvature tests.

Acknowledgments

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Applied Mathematics program under Contract No. DE-AC02-06CH11357. Victor M. Zavala acknowledges funding from the DOE Office of Science under the Early Career program. We also acknowledge the computing resources provided by the Laboratory Computing Resource Center at Argonne National Laboratory.

References

[1] N. Arora and L. T. Biegler. A trust region sqp algorithm for equality constrained parameter estimation with simple parameter bounds. *Computational Optimization and Applications*, 28(1):51–86, 2004.

[2] Y. Bard. *Nonlinear Parameter Estimation*. Academic Press, Cambridge, MA, 1974.

[3] G. Biros and O. Ghattas. Parallel Lagrange–Newton–Krylov–Schur methods for PDE-constrained optimization. Part I: The Krylov–Schur solver. *SIAM Journal on Scientific Computing*, 27(2):687–713, 2005.

[4] C. Boehm, R. Findeisen, and F. Allgoewer. Avoidance of Poorly Observable Trajectories: A predictive control perspective. In *Proceedings of the 17th IFAC World Congress, Seoul, Korea*, pages 1952–1957, 2008.

[5] J. Bunch and L. Kaufman. Some stable methods for calculating inertia and solving symmetric linear systems. *Mathematics of computation*, pages 163–179, 1977.

[6] R. H. Byrd, J. C. Gilbert, and J. Nocedal. A trust-region method based on interior-point techniques for nonlinear programming. *Math. Programm.*, 89:149–185, 2000.

[7] N. Chiang, C. G. Petra, and V. M. Zavala. Structured nonconvex optimization of large-scale energy systems using PIPS-NLP. In *Proc. of the 18th Power Systems Computation Conference (PSCC), Wroclaw, Poland*, 2014.

[8] N.-Y. Chiang and V. M. Zavala. An inertia-free filter line-search algorithm for large-scale nonlinear programming. *Computational Optimization and Applications*, To Appear, 2015.

[9] A. R. Conn, N. I. Gould, and P. L. Toint. *Trust region methods*, volume 1. Siam, 2000.

[10] F. Curtis, O. Schenk, and A. Wächter. An interior-point algorithm for large-scale nonlinear optimization with inexact step computations. *SIAM Journal on Scientific Computing*, 32(6):3447–3475, 2010.

[11] M. Diehl, R. Amrit, and J. B. Rawlings. A lyapunov function for economic optimizing model predictive control. *Automatic Control, IEEE Transactions on*, 56(3):703–707, 2011.

[12] I. S. Duff. Ma57—a code for the solution of sparse symmetric definite and indefinite systems. *ACM Transactions on Mathematical Software (TOMS)*, 30(2):118–144, 2004.

[13] R. Fourer, D. M. Gay, and B. W. Kernighan. *AMPL*. Boyd and Fraser, 1993.

[14] N. I. Gould, M. E. Hribar, and J. Nocedal. On the solution of equality constrained quadratic programming problems arising in optimization. *SIAM Journal on Scientific Computing*, 23(4):1376–1395, 2001.

[15] E. L. Haseltine and J. B. Rawlings. Critical evaluation of extended kalman filtering and moving-horizon estimation. *Industrial & engineering chemistry research*, 44(8):2451–2460, 2005.

[16] M. Heinkenschloss and D. Ridzal. A matrix-free trust-region sqp method for equality constrained optimization. *SIAM Journal on Optimization*, 24(3):1507–1541, 2014.

[17] R. Huang, E. Harinath, and L. T. Biegler. Lyapunov stability of economically oriented nmpc for cyclic processes. *Journal of Process Control*, 21(4):501–509, 2011.

[18] J. Kang, Y. Cao, D. P. Word, and C. Laird. An interior-point method for efficient solution of block-structured nlp problems using an implicit schur-complement decomposition. *Computers & Chemical Engineering*, 71:563–573, 2014.

[19] J. Kang, N. Chiang, C. D. Laird, and V. M. Zavala. Nonlinear programming strategies on high-performance computers. In *Proc. of the IEEE Conference on Decision and Control, Osaka, Japan*, 2015.

[20] G. Karypis and V. Kumar. Metis-unstructured graph partitioning and sparse matrix ordering system, version 2.0. 1995.

[21] S. Mumma. Transient occupancy ventilation by monitoring co_2. *ASHRAE IAQ Applications*, pages 21–23, 2004.

[22] J. Nocedal and S. Wright. *Numerical Optimization*. Springer, New York, NY, 1999.

[23] C. V. Rao, J. B. Rawlings, and D. Q. Mayne. Constrained state estimation for nonlinear discrete-time systems: Stability and moving horizon approximations. *IEEE Trans. Automat. Contr.*, 48:246–258, 2003.

[24] M. Santamouris, A. Synnefa, M. Asssimakopoulos, I. Livada, K. Pavlou, M. Papaglastra, N. Gaitani, D. Kolokotsa, and V. Assimakopoulos. Experimental investigation of the air flow and indoor carbon dioxide concentration in classrooms with intermittent natural ventilation. *Energy and Buildings*, 40(10):1833–1843, 2008.

[25] O. Schenk, A. Wächter, and M. Hagemann. Matching-based preprocessing algorithms to the solution of saddle-point problems in large-scale nonconvex interior-point optimization. *Computational Optimization and Applications*, 36:321–341, 2007.

[26] I.-B. Tjoa and L. Biegler. Reduced successive quadratic programming strategy for errors-in-variables estimation. *Computers & chemical engineering*, 16(6):523–533, 1992.

[27] A. Wächter and L. Biegler. Line search filter methods for nonlinear programming: Motivation and global convergence. *SIAM Journal on Optimization*, 16(1):1–31, 2005.

[28] A. Wächter and L. T. Biegler. On the implementation of a primal-dual interior point filter line search algorithm for large-scale nonlinear programming. *Mathematical Programming*, 106:25–57, 2006.

[29] R. A. Waltz, J. L. Morales, J. Nocedal, and D. Orban. An interior algorithm for nonlinear optimization that combines line search and trust region steps. *Mathematical Programming*, 107(3):391–408, 2006.

[30] S. Wang, J. Burnett, and H. Chong. Experimental validation of co2-based occupancy detection for demand-controlled ventilation. *Indoor and Built Environment*, 8(6):377–391, 1999.

[31] V. M. Zavala. Stability analysis of an approximate scheme for moving horizon estimation. *Computers & Chemical Engineering*, 34(10):1662–1670, 2010.

[32] V. M. Zavala. Inference of building occupancy signals using moving horizon estimation and fourier regularization. *Journal of Process Control*, 24(6):714–722, 2014.

[33] V. M. Zavala. Stochastic optimal control model for natural gas networks. *Computers & Chemical Engineering*, 64:103–113, 2014.

[34] V. M. Zavala. A multiobjective optimization perspective on the stability of economic MPC. *IFAC-PapersOnLine*, 48(8):974 – 980, 2015. 9th IFAC Symposium on Advanced Control of Chemical Processes ADCHEM. Whistler, Canada.

[35] V. M. Zavala and L. T. Biegler. Large-scale parameter estimation in low-density polyethylene tubular reactors. *Ind. Eng. Chem. Res.*, 45:7867–7881, 2006.

[36] V. M. Zavala and L. T. Biegler. Nonlinear programming strategies for state estimation and model predictive control. In *Nonlinear model predictive control*, pages 419–432. Springer, 2009.

[37] V. M. Zavala and L. T. Biegler. Optimization-based strategies for the operation of low-density polyethylene tubular reactors: nonlinear model predictive control. *Computers & Chemical Engineering*, 33(10):1735–1746, 2009.