Janick Frasch+

Dual Newton Strategies for Model Predictive Control

University of Freiburg
October 23, 2014

+ based on joint work with Dominik Alfke, Moritz Diehl, Joachim Ferreau, Boris Houska, Sebastian Sager, Milan Vukov, ...

Faculty of Mathematics
Otto-von-Guericke University Magdeburg, Germany

Department of Electrical Engineering
KU Leuven, Belgium

Motivation

[Source: http://www.guenstige-risikolebensversicherung.de/]

Motivation

[Source: http://www.guenstige-risikolebensversicherung.de/]

(Some) requirements for good control and estimation performance:

- long prediction horizons
- short reaction times $\rightarrow$ fast algorithms

Solution framework for dynamic optimization

Discretized OCP

$$\begin{aligned} \min_{x,u} \quad & \sum_{k=0}^{N} \ell_k(x_k, u_k) \\ \text{s.t.} \quad & x_{k+1} = F_k(x_k, u_k) \qquad \forall k = 0, \dots, N-1 \\ & x_0 = \hat{x}_0 \\ & 0 \le r_k(x_k, u_k) \qquad \forall k = 0, \dots, N \end{aligned}$$

$\Downarrow$

Linearization

$\Downarrow$

Highly structured QP

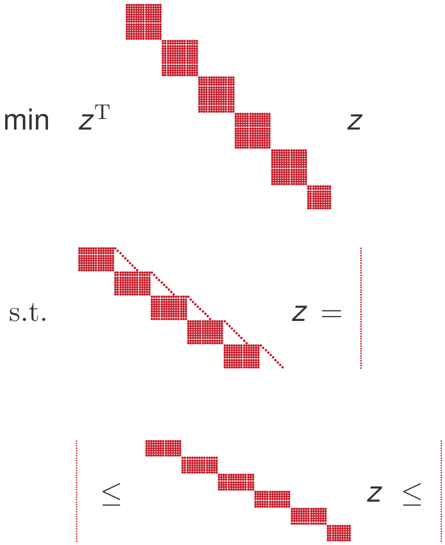
$$\begin{aligned} \min_{z} \quad & \sum_{k=0}^{N} \left(\frac{1}{2} z_k^{\mathrm{T}} H_k z_k + g_k^{\mathrm{T}} z_k \right) \\ \text{s.t.} \quad & E_{k+1} z_{k+1} = C_k z_k + c_k \qquad \forall k = 0, \dots, N-1 \\ & \underline{d}_k \le D_k z_k \le \overline{d}_k \qquad \forall k = 0, \dots, N \end{aligned}$$

- $x_k \in \mathbb{R}^{n_x}$ system state
- $u_k \in \mathbb{R}^{n_u}$ control inputs
- $x_0 \in \mathbb{R}^{n_x}$ initial value
- F_k hides IVP solution
- ℓ_k, F_k, r_k possibly nonlinear

Linearization schemes:

- SQP: iterative linearization and QP solution
- RTI: one QP per sampling time
- MLI: partial and/or inexact linearizations

QP sparsity patterns in dynamic optimization



QP sparsity exploitation I: Condensing [BP1984,AFVD13,...]

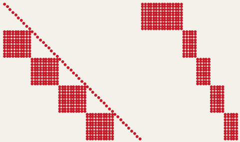
Split variables

Partitioning: $v = (x_1, ..., x_N)$, $w = (x_0, u_0, ..., u_{N-1})$

$$\min_{v,w} \ \frac{1}{2}\begin{bmatrix} v \\ w \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} H_{vv} & H_{wv} \\ H_{vw} & H_{ww} \end{bmatrix} \begin{bmatrix} v \\ w \end{bmatrix} + \begin{bmatrix} g_v \\ g_w \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} v \\ w \end{bmatrix}$$

$$\text{s.t.} \quad 0 = C_v v + C_w w + c$$

$$\underline{d} \le D \begin{bmatrix} v \\ w \end{bmatrix} \le \overline{d}$$



Eliminate $v := C_v^{-1} c + C_v^{-1} C_w w$

Solve reduced-size, dense QP

$$\min_{w} \ \frac{1}{2} w^{\mathrm{T}} H_{\text{cond}} w + g_{\text{cond}}^{\mathrm{T}} w$$

$$\text{s.t.} \quad \underline{d}_{\text{cond}} \le D_{\text{cond}} w \le \overline{d}_{\text{cond}}$$

Properties:

- Elimination quadratic in horizon length
- Dense QP of size $N\, n_u$

QP sparsity exploitation II: Interior point methods

Highly structured QP

$$\min_{z,s} \quad \sum_{k=0}^{N} \left(\frac{1}{2} \, z_k^{\mathrm{T}} H_k z_k + g_k^{\mathrm{T}} z_k \right)$$

$$\text{s.t.} \quad E_{k+1} z_{k+1} = C_k z_k + c_k \qquad \forall\, k = 0, \ldots, N-1$$

$$0 = D_k z_k - d_k + s_k \qquad \forall\, k = 0, \ldots, N$$

Linearize KKT system

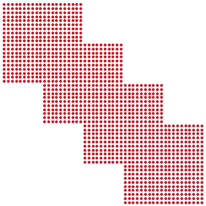
$$\begin{bmatrix} \mathcal{H} & \mathcal{C}^{\mathrm{T}} & \mathcal{D}^{\mathrm{T}} & \\ \mathcal{C} & & & \\ \mathcal{D} & & & I \\ & & \mathcal{S} & \mathcal{M} \end{bmatrix} \begin{bmatrix} \Delta z \\ \Delta \lambda \\ \Delta \mu \\ \Delta s \end{bmatrix} = - \begin{bmatrix} r_{\mathcal{L}} \\ r_{\mathrm{eq}} \\ r_{\mathrm{ieq}} \\ r_s \end{bmatrix}$$

Perform Newton steps

$$\begin{bmatrix} z & \lambda & \mu & s \end{bmatrix} += \alpha \begin{bmatrix} \Delta z & \Delta \lambda & \Delta \mu & \Delta s \end{bmatrix}$$

Characteristics:

- Sparsity of linear system:



- Choice of right-hand side depends on specific method (e.g., barrier parameter)
- Tailored factorization possible
- Does not benefit from similarity between problems (“warmstarting”)

Exploiting QP structure III: Dual decomposition

Highly structured QP

$$\begin{aligned} \min_{z} \quad & \sum_{k=0}^{N} \left(\frac{1}{2} z_k^{\mathrm{T}} H_k z_k + g_k^{\mathrm{T}} z_k \right) \\ \text{s.t.} \quad & E_{k+1} z_{k+1} = C_k z_k + c_k \qquad \forall\, k = 0, \ldots, N-1 \\ & \underline{d}_k \leq D_k z_k \leq \overline{d}_k \qquad \forall\, k = 0, \ldots, N \end{aligned}$$

Assumptions

- $H_k \succ 0$
- feasible

Partial dualization

$$\max_{\lambda} \min_{z} \sum_{k=0}^{N} \left(\frac{1}{2} z_k^{\mathrm{T}} H_k z_k + g_k^{\mathrm{T}} z_k \right) + \sum_{k=0}^{N-1} \lambda_{k+1}^{\mathrm{T}} \left(C_k z_k + c_k - E_{k+1} z_{k+1} \right)$$

$$\text{s.t.} \quad \underline{d}_k \leq D_k z_k \leq \overline{d}_k \qquad \forall k = 0, \ldots, N$$

Separable dual function

$$\max_{\lambda} \min_{z} \sum_{k=0}^{N} \left(\frac{1}{2} z_k^{\mathrm{T}} H_k z_k + \left(g_k^{\mathrm{T}} + \begin{bmatrix} \lambda_k \\ \lambda_{k+1} \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} -E_k \\ C_k \end{bmatrix} \right) z_k + \lambda_{k+1}^{\mathrm{T}} c_k \right)$$

$$\text{s.t.} \quad \underline{d}_k \leq D_k z_k \leq \overline{d}_k \qquad \forall k = 0, \ldots, N$$

A separable two-level reformulation

Unconstrained consensus problem

$$\max_{\lambda} f^*(\lambda) := \sum_{k=0}^{N} f_k^*(\lambda)$$

Parametric stage problems

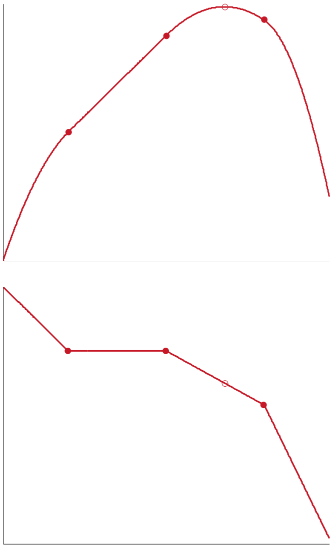
$$f_k^*(\lambda) := \min_{z_k} \frac{1}{2} z_k^{\mathrm{T}} H_k z_k + p_k(\lambda)^{\mathrm{T}} z_k + q_k(\lambda)$$

$$\text{s.t. } \underline{d}_k \leq D_k z_k \leq \overline{d}_k,$$

Properties of f^*

- concave
- piecewise quadratic

  ($z^*(\lambda)$ continuous, piecewise affine [Fiacco83, Zafiriou90])
- $f^* \in C^1$ [e.g., Bertsekas1997]
- $\frac{\partial^2 f^*}{\partial \lambda^2}(\lambda)$ constant within each primal active set



Dual (nonsmooth) Newton strategy

- Unconstrained concave high-level problem

$$\max_{\lambda} f^*(\lambda)$$

- Apply Newton's method

$$\lambda^{i+1} := \lambda^i + \alpha \Delta\lambda$$

where

$$\left[\frac{\partial^2 f^*}{\partial \lambda^2}(\lambda^i) \right] \Delta\lambda = - \left[\frac{\partial f^*}{\partial \lambda}(\lambda^i) \right]$$

- Globalization needed due to kinks
- Convergence under mild assumptions [Frasch,Sager&Diehl 2014 (submitted); related proofs in: Qi&Sun 1993, Li&Swetits 1997]

Sparsity patterns of the Newton system

Dual function

$$\sum_{k=0}^{N} \quad f_k^*(\lambda) = \min_{z_k} \frac{1}{2} z_k^{\mathrm{T}} H_k z_k + p_k(\lambda_k, \lambda_{k+1})^{\mathrm{T}} z_k + q_k(\lambda_k, \lambda_{k+1})$$

$$\text{s.t. } \underline{d}_k \leq D_k z_k \leq \overline{d}_k,$$

Structure of the Newton system

$$\begin{bmatrix} \frac{\partial^2 f^*}{\partial \lambda_1^2} & \frac{\partial^2 f^*}{\partial \lambda_1 \lambda_2} & & \\ \frac{\partial^2 f^*}{\partial \lambda_2 \lambda_1} & \frac{\partial^2 f^*}{\partial \lambda_2^2} & \ddots & \\ & \ddots & \ddots & \frac{\partial^2 f^*}{\partial \lambda_{N-1} \lambda_N} \\ & & \frac{\partial^2 f^*}{\partial \lambda_N \lambda_{N-1}} & \frac{\partial^2 f^*}{\partial \lambda_N^2} \end{bmatrix} \begin{bmatrix} \Delta\lambda_1 \\ \Delta\lambda_2 \\ \vdots \\ \Delta\lambda_N \end{bmatrix} = \begin{bmatrix} \frac{\partial f_0^*}{\partial \lambda_1} + \frac{\partial f_1^*}{\partial \lambda_1} \\ \frac{\partial f_1^*}{\partial \lambda_2} + \frac{\partial f_2^*}{\partial \lambda_2} \\ \vdots \\ \frac{\partial f_{N-1}^*}{\partial \lambda_N} + \frac{\partial f_N^*}{\partial \lambda_N} \end{bmatrix}$$

→ Tailored Cholesky factorization

Solution of stage QPs

Dual function

$$\sum_{k=0}^{N} \qquad f_k^*(\lambda) = \min_{z_k} \frac{1}{2} z_k^{\mathrm{T}} H_k z_k + \left(g_k^{\mathrm{T}} + \begin{bmatrix} \lambda_k \\ \lambda_{k+1} \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} -E_k \\ C_k \end{bmatrix} \right) z_k + \lambda_{k+1}^{\mathrm{T}} c_k$$

$$\text{s.t. } \underline{d}_k \leq D_k z_k \leq \overline{d}_k,$$

Stage QP

$$f_k^*(\lambda) := \min_{z_k} \frac{1}{2} z_k^{\mathrm{T}} H_k z_k + p_k(\lambda)^{\mathrm{T}} z_k + q_k(\lambda)$$

$$\text{s.t. } \underline{d}_k \leq D_k z_k \leq \overline{d}_k,$$

- Parametric gradient, Hessian constant
- General case: parametric active set strategy (e.g., qpOASES [Ferreau et. al, 2008, 2014])
- diagonal H, identity D: clipping

$$z_k^* := \max(\underline{d}_k, \min(z_k, \overline{d}_k))$$

Dual problem setup

Dual function

$$\sum_{k=0}^{N} \quad f_k^*(\lambda) = \min_{z_k} \frac{1}{2} z_k^{\mathrm{T}} H_k z_k + \left(g_k^{\mathrm{T}} + \begin{bmatrix} \lambda_k \\ \lambda_{k+1} \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} -E_k \\ C_k \end{bmatrix} \right) z_k + \lambda_{k+1}^{\mathrm{T}} c_k$$

$$\text{s.t. } \underline{d}_k \leq D_k z_k \ \leq \ \overline{d}_k,$$

Dual gradient

$$\begin{bmatrix} \frac{\partial f_k^*}{\partial \lambda_k} \\ \frac{\partial f_k^*}{\partial \lambda_{k+1}} \end{bmatrix} = - \left(\begin{bmatrix} -E_k \\ C_k \end{bmatrix} z_k^* + \begin{bmatrix} 0 \\ c_k \end{bmatrix} \right)$$

- Analytical, closed-form expressions
- Concurrent, block-wise setup possible

Hessian blocks

$$\frac{\partial^2 f^*}{\partial \lambda_k \lambda_{k+1}} = E_{k+1} \underbrace{\frac{\partial z_{k+1}^*}{\partial \lambda_k}}_{=0} - C_k \frac{\partial z_k^*}{\partial \lambda_k} = -C_k P_k^* E_k^{\mathrm{T}}$$

$$\frac{\partial^2 f^*}{\partial \lambda_k \lambda_k} = E_k \frac{\partial z_k^*}{\partial \lambda_k} - C_{k-1} \frac{\partial z_{k-1}^*}{\partial \lambda_k} = E_k P_k^* E_k^{\mathrm{T}} + C_{k-1} P_{k-1}^* C_{k-1}^{\mathrm{T}}$$

- Constraint Nullspace elimination matrix $P_k^* := Z_k^* \left(Z_k^{*\mathrm{T}} H_k Z_k^*\right)^{-1} Z_k^{*\mathrm{T}}$
- Nullspace basis matrix Z_k^* of stage problem

Bottom-up Hessian factorization

Observation

- Hessian blocks change only if $\{C_k, E_k\} \left(Z_k^* \, (Z_k^{*\,\mathrm{T}} H_k \, Z_k^*)^{-1} Z_k^{*\,\mathrm{T}} \right) \{C_k, E_k\}^{\mathrm{T}}$ changes
- Change triggered by active-set change of stage QP

Assumption

- Few active-set changes on last intervals
- Motivation: tracking MPC problems, LQR terminal cost

Implications for factorization

- Invert elimination order in Cholesky factorization (“backwards in time”)
- Start factorization only at last stage with active set change
- Better numerical stability in practice (singular Hessian caused by active constraints)

Parallelization aspects

Separable dual function

$$\sum_{k=0}^{N} \qquad f_k^*(\lambda) = \min_{z_k} \frac{1}{2} z_k^{\mathrm{T}} H_k z_k + \left(g_k^{\mathrm{T}} + \begin{bmatrix} \lambda_k \\ \lambda_{k+1} \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} -E_k \\ C_k \end{bmatrix} \right) z_k + \lambda_{k+1}^{\mathrm{T}} c_k$$

$$\text{s.t.} \ \underline{d}_k \leq D_k z_k \ \leq \ \overline{d}_k,$$

- Concurrent evaluation of stage QPs possible
- Setup of Newton system parallelizable

- $\log N$ parallel Hessian factorization via cyclic reduction [Wright, 1991]

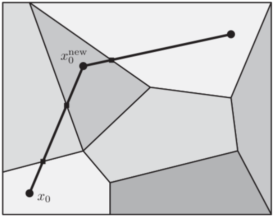
Exact line search

Line search

- Goal: find $\alpha^i \approx \arg\max_{0 \leq \alpha \leq 1} f^*(\lambda^i + \alpha \Delta\lambda^i)$

Parametric active set strategy [Ferreau et al., 2008]

- tracks $z_k^*(\lambda^i + \tau\Delta\lambda^i)$ for $\tau \in [0, 1]$
- $f_k^*(\tau)$ is univariate piecewise quadratic concave spline.



[Ferreau et al., 2008]

Exact line search

- logging of spline base points in parametric active set strategy
- superpositioning over stages

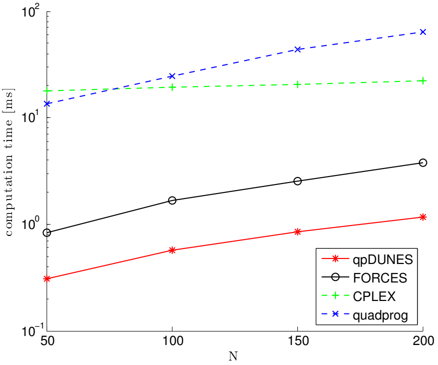
Software implementation

qpDUNES — An implementation of the *DU*al *NE*wton *S*trategy

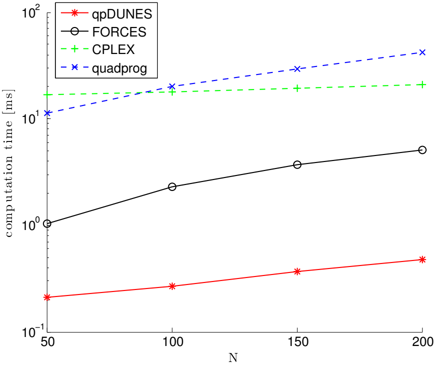
- Open-source sparse QP solver
- Plain ANSI C
- Custom linear algebra
- Dynamic memory for flexibility, static for performance (soon :)
- Linear MPC from C/C++ and Matlab
- Usable as sparse QP solver within ACADO Toolkit [Houska et al. 2009, 2011]
  - Nonlinear MPC
  - Moving Horizon Estimation
- Version with support for affine constraints not yet public

`http://mathopt.de/qpDUNES`

Linear MPC Benchmarking: Double Integrator

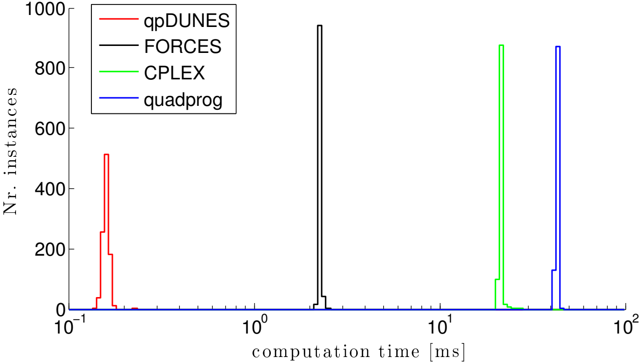


Cold started

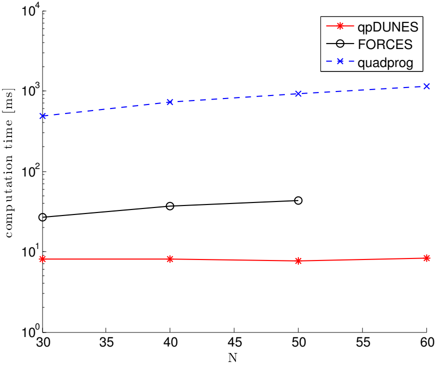


Warm started

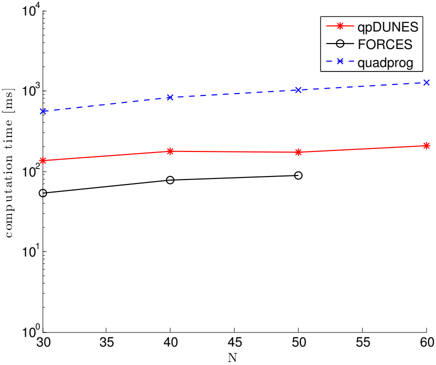
Linear MPC: Oscillating masses from [Wang & Boyd, 2010]



Hanging chain of masses: linear MPC ($M = 5$)

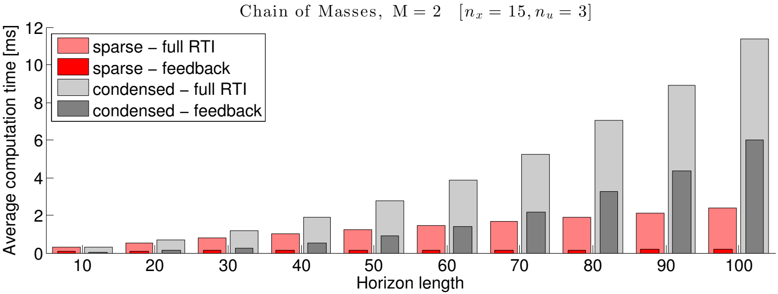


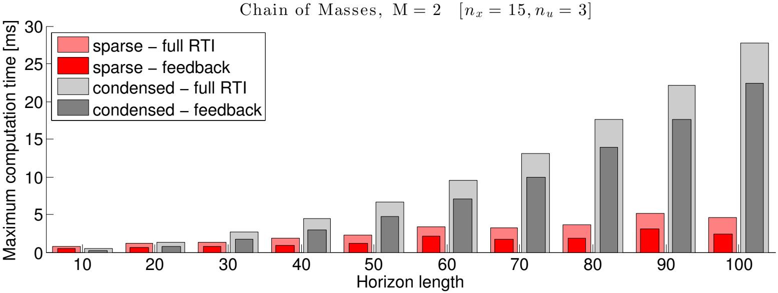
Mean computation times



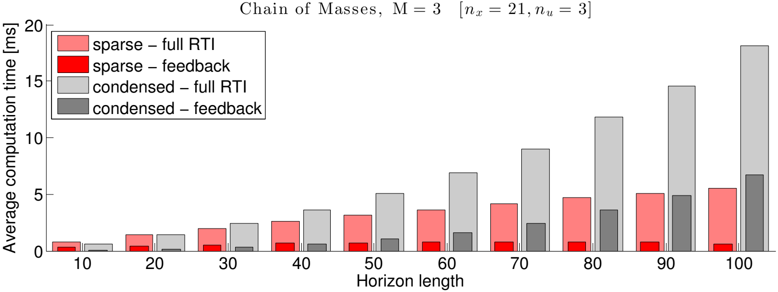
Maximum computation times

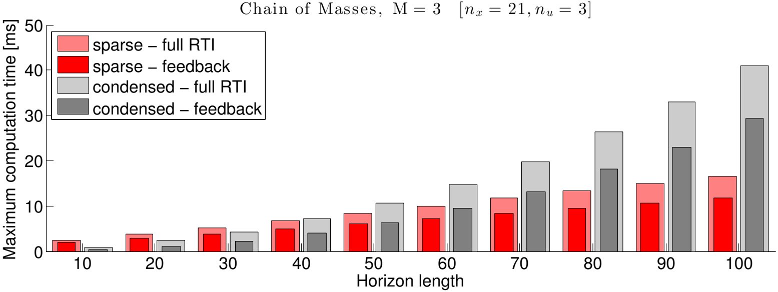
Hanging chain of masses: nonlinear MPC ($M = 2$)



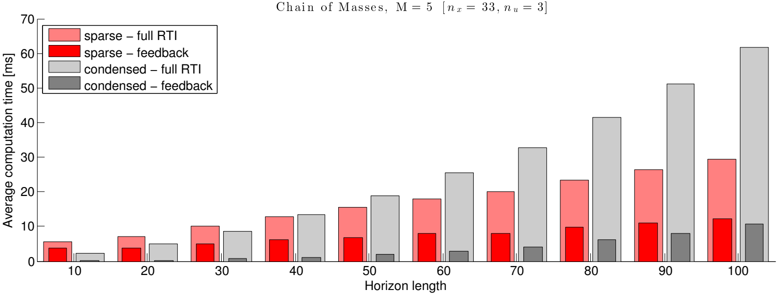


Hanging chain of masses: nonlinear MPC ($M = 3$)





Challenges to qpDUNES' performance



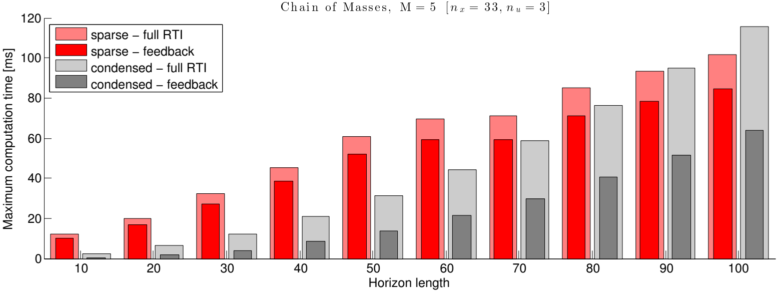
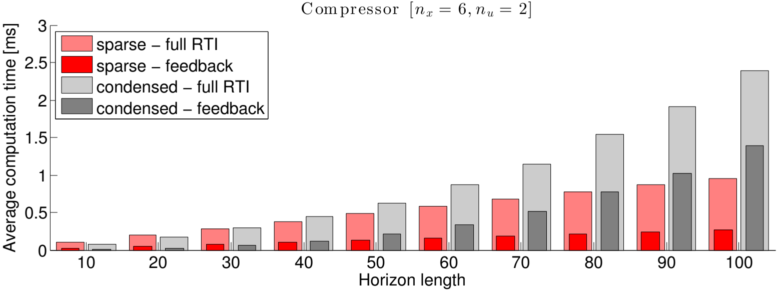
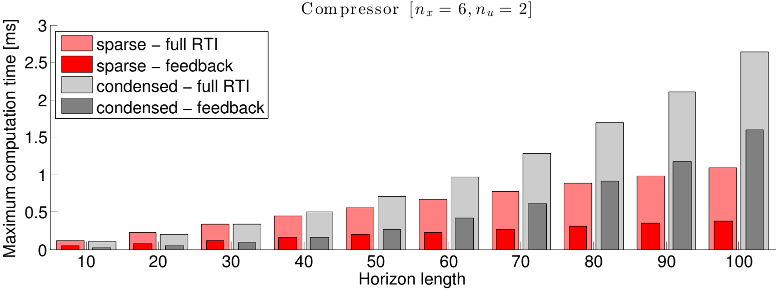
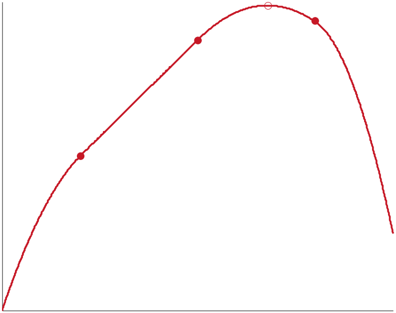


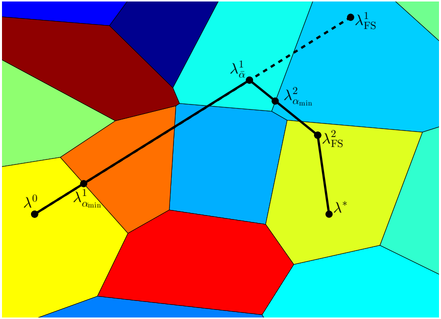
ABB Compressor: nonlinear MPC





Performance analysis





Comparison with dual decomposition first-order methods

- DUNES requires drastically fewer iterations ($\approx 10^0 - 10^1$)
- DUNES exploits dual function curvature to predict active set

Warmstarting of DUNES

Guaranteed active set change

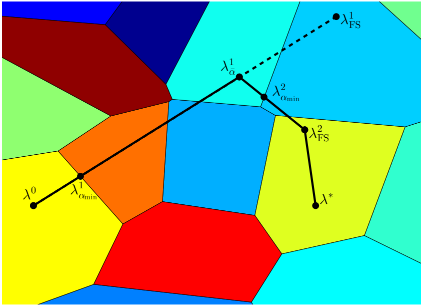
- If Newton Hessian unregularized
- Intrinsic due to piecewise quadratic nature
- Possibly many active set changes per iteration

Shifting policy

$$\lambda_k^0 := \lambda_{k+1}^* \qquad \forall\, k = 1, \ldots, N-1$$
$$\lambda_N^0 := \lambda_N^*$$

One-step terminal convergence

- f^* quadratic within each primal AS
- Newton's method finds quadratic minimizer
- Nominal MPC: convergence in first iteration (even NMPC)



Loss of Hessian regularity

- Want to solve Newton system $M(\lambda)\,\Delta\lambda = G(\lambda)$
- Dual Hessian given by $M(\lambda) = C\,Z^* \left(Z^{*\mathrm{T}} H\,Z^*\right)^{-1} Z^{*\mathrm{T}} C^{\mathrm{T}}$
- For non-optimal iterates (despite LICQ):
  Active stage constraints may be linearly dependent across stages
- Dual Hessian may turn out singular

Regularization

Solve modified Newton system:

$$(M(\lambda) + \gamma\, I\,)\,\Delta\lambda = G(\lambda)$$

- Large γ improves conditioning
- but: perturbs Newton step

Preconditioned gradient steps

- Default to $\bar{M} = C\,H^{-1} C^{\mathrm{T}}$
- $\bar{M} \succ 0$ by LICQ
- Solve $\bar{M}\,\Delta\lambda = G(\lambda)$
- Can show $\bar{M} \succeq M(\lambda)$ for any λ
- No globalization needed
- $\bar{M}$ can be factorized offline

A DUNES for Non-Convex Optimization [Houska, F., Diehl 2014]

Linearly coupled NLP

$$\min_{z} \quad \sum_{k=0}^{N} \ell_k(z_k)$$

$$\text{s.t.} \quad \sum_{k=0}^{N} C_k z_k = c_k$$

$$d_k(z_k) \leq 0 \qquad \forall\, k = 0, \dots, N$$

Here: ℓ_k, d_k possibly non-convex

Hitches:

- Non-zero duality gap
- Non-unique multipliers
- Loss of original convergence guarantees

Original stage problem

$$\min_{z_k} \quad \ell_k(z_k) + \lambda^{\mathrm{T}} C_k\, z_k$$

$$\text{s.t.} \quad d_k(z_k) \leq 0$$

A DUNES for Non-Convex Optimization [Houska, F., Diehl 2014]

Linearly coupled NLP

$$\min_{z} \sum_{k=0}^{N} \ell_k(z_k)$$

$$\text{s.t.} \quad \sum_{k=0}^{N} C_k z_k = c_k$$

$$d_k(z_k) \leq 0 \qquad \forall\, k = 0, \ldots, N$$

Here: ℓ_k, d_k possibly non-convex

Hitches:

- Non-zero duality gap
- Non-unique multipliers
- Loss of original convergence guarantees

Remedy: use augmented-Lagrangian stage problem formulation

Augmented stage problem

$$\min_{z_k} \quad \ell_k(z_k) + \lambda^{\mathrm{T}} C_k\, z_k + \frac{\rho}{2} \| z_k - y_k \|_{\Sigma_k}^2$$

$$\text{s.t.} \quad d_k(z_k) \leq 0$$

- ADMM-inspired
- λ updated from dual consensus problem
- y_k primal variables “associated” with λ
- convergence requires $\Sigma_k \succ 0$, ρ sufficiently large

qpDUNES roadmap

Current status

- Open source software
- linear MPC interfaces from C/C++ and Matlab
- available for nonlinear MPC in ACADO
- diagonal H_k, simple bounds: public
- affine constraints: group internal

Open software issues

- parallelization
- code generation and static memory version
- efficient cold-starting
- Support for collocation problems?
- Support for non-convex/not strictly convex problems?

References and further reading

J. V. Frasch, M. Vukov, H. J. Ferreau, M. Diehl. *A new quadratic programming strategy for efficient sparsity exploitation in SQP-based nonlinear MPC and MHE*. Proceedings of the 19th IFAC World Congress, 2014.

J. V. Frasch, S. Sager, M. Diehl. *A Parallel Quadratic Programming Method for Dynamic Optimization Problems*. 201x (under review). `http://www.optimization-online.org/DB_HTML/2013/11/4114.html`

B. Houska, J. V. Frasch, M. Diehl. *An Augmented Lagrangian based Algorithm for Distributed Non-Convex Optimization*. SIAM Journal on Optimization, 201x (under review). `http://www.optimization-online.org/DB_FILE/2014/07/4427.html`

ACADO Toolkit Homepage. `http://www.acadotoolkit.org`

qpDUNES Homepage. `http://mathopt.de/qpDUNES`

H.J. Ferreau, H. G. Bock, M. Diehl. *An online active set strategy to overcome the limitations of explicit MPC*. International Journal of Robust and Nonlinear Control 18(8):816-830, 2008.

H.J. Ferreau, A. Kozma, M. Diehl. *A parallel active-set strategy to solve sparse parametric quadratic programs arising in MPC*. Proceedings of the 4th IFAC Nonlinear Model Predictive Control Conference, 2012.

W. Li and J. Swetits. *A new algorithm for solving strictly convex quadratic programs*. SIAM Journal of Optimization, 7(3):595–619, 1997.

M. Vukov, A. Domahidi, H. J. Ferreau, M. Morari and M. Diehl. *Auto-generated Algorithms for Nonlinear Model Predicitive Control on Long and on Short Horizons*. Proceedings of the 52nd Conference on Decision and Control (CDC), 2013.

S. Wright. *Partitioned dynamic programming for optimal control*. SIAM Journal on Optimization, 1(4):620-642, 1991.

http://mathopt.de/qpDUNES/

Janick Frasch

Institute of Mathematical Optimization (IMO) — Room 2.23
Faculty of Mathematics
Otto-von-Guericke University Magdeburg
Universitaetsplatz 2
D-39106 Magdeburg, Germany

E-Mail: frasch@ovgu.de
Phone: +49 391 67-11450