Janick Frasch+

qpDUNES — a dual Newton strategy for convex QP

Workshop "Embedded Quadratic Programming", Freiburg
March 19, 2014

+ joint work with Moritz Diehl, Joachim Ferreau, Attila Kozma, Sebastian Sager, Milan Vukov

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

Department of Electrical Engineering
KU Leuven, Belgium

Nonlinear model predictive control (MPC)

Discretized OCP

$$\min_{x,u} \quad \sum_{k=0}^{N} \ell_k\left(x_k, u_k\right)$$

$$\text{s.t.} \quad x_{k+1} = F_k\left(x_k, u_k\right) \qquad \forall\, k = 0, \ldots, N-1$$

$$x_0 = p$$

$$0 \leq r_k\left(x_k, u_k\right) \qquad \forall\, k = 0, \ldots, N$$

$\Downarrow$

Sequential quadratic programming (SQP)

$\Downarrow$

Highly structured QP

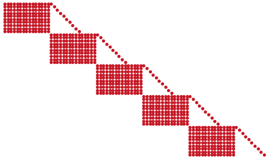
$$\min_{z} \quad \sum_{k=0}^{N} \left(\frac{1}{2}\; z_k^{\mathrm{T}} H_k z_k + g_k^{\mathrm{T}} z_k \right)$$

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

$$\underline{d}_k \leq D_k z_k \leq \overline{d}_k \qquad \forall\, k = 0, \ldots, N$$

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

- RTI: only one QP per sampling time



Exploiting QP structure I: Condensing

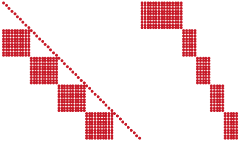
Condensing of the sparse QP

Partitioning $v := \left[\Delta x_1, ..., \Delta x_N\right]$, $w := \left[\Delta x_0, \Delta u_0, ..., \Delta u_{N-1}\right]$:

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

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

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



- Eliminate $v := C_v^{-1} c + C_v^{-1} C_w w$
- Solve smaller QP in w with dense solver

Condensed QP

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

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

Drawbacks:

- requires expensive condensing step
- dense QP of size $N n_u$

Exploiting QP structure II: Interior Point methods

Highly structured QP

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

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

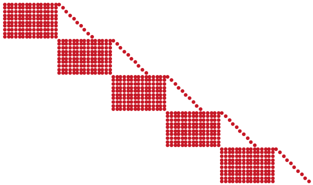
$$0 = D_k z_k - d_k + s_k \qquad \forall\, k = 0, \ldots, N$$

Linearize KKT system

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

Perform Newton steps

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



- Choice of right-hand side depends on specific method (e.g., barrier parameter)
- Tailored factorization possible

Drawback:

- Cannot exploit similarity between problems (“warmstarting”)

Exploiting QP structure III: Dual Decomposition

Highly structured QP

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

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

$$\underline{d}_k \leq D_k z_k \leq \overline{d}_k \qquad \forall\, k = 0, \ldots, N$$

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

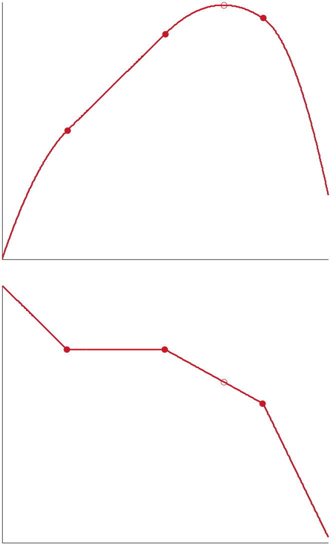
$$f_k^*(\lambda) := \min_{z_k} \frac{1}{2} z_k^{\mathrm{T}} H_k z_k + p_k(\lambda)^{\mathrm{T}} z_k + q_k(\lambda)$$

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

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]

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

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

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
- Newton Hessian might be indefinite
- Levenberg-Marquardt or “on-the-fly” regularization

Analytical gradient computation

Dual function

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

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

Dual gradient

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

- $\frac{\partial f^*}{\partial z} \frac{\partial z}{\partial \lambda}$ terms vanish
- follows from Danskin's theorem [e.g., Bersekas 1997]
- easy to see via chain rule and stationarity property

Analytical Hessian computation

Dual gradient

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

Hessian blocks

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

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

- Constraint Nullspace elimination matrix
  $P_k^* := Z_k^* ({Z_k^*}^{\mathrm{T}} H_k Z_k^*)^{-1} {Z_k^*}^{\mathrm{T}} \in \mathbb{R}^{n_z \times n_z}$
- Nullspace basis matrix Z_k^* of $\mathrm{QP}_k \in \mathbb{R}^{n_z \times (n_z - n_{\text{act}})}$
- Z_k^* and symmetric factor of $({Z_k^*}^{\mathrm{T}} H_k Z_k^*)^{-1}$ often for free in nullspace method

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)

Warmstarting of the dual Newton strategy

Guaranteed active set change

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

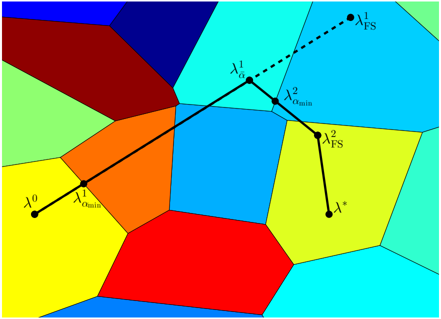
Shifting policy

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

1-step terminal convergence

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

Steps of the dual Newton strategy:



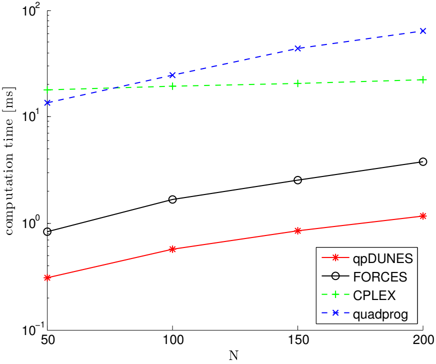
Software implementation

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

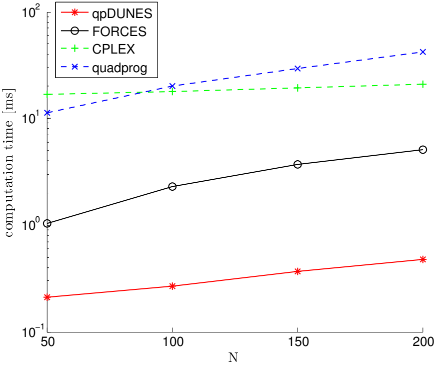
- Open-source sparse QP solver
- Plain ANSI C
- Custom linear algebra
- Dynamic memory for flexibility, static for performance (soon :)
- Linear MPC from C/C++ and Matlab
- Usable as sparse QP solver within ACADO Toolkit [Houska et al. 2009, 2011]
  - Nonlinear MPC
  - Moving Horizon Estimation
- Version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

http://mathopt.de/qpDUNES

Linear MPC Benchmarking: Double Integrator

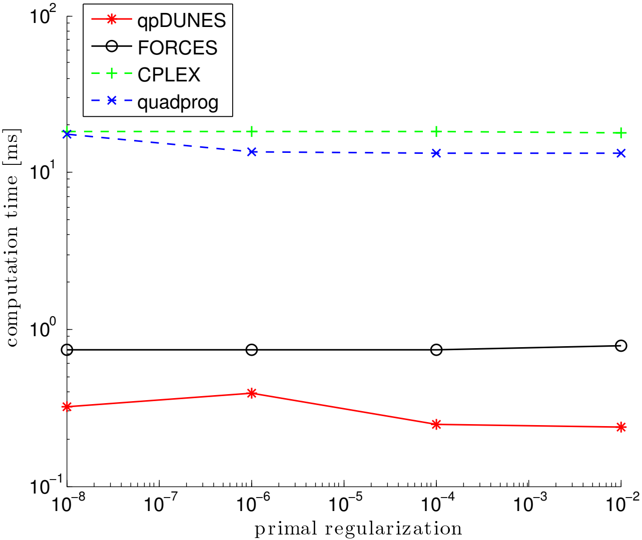


Cold started

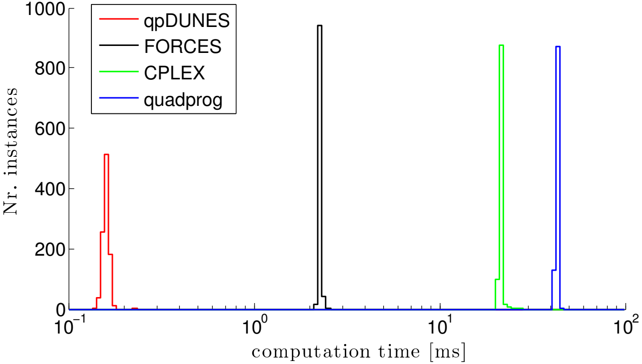


Warm started

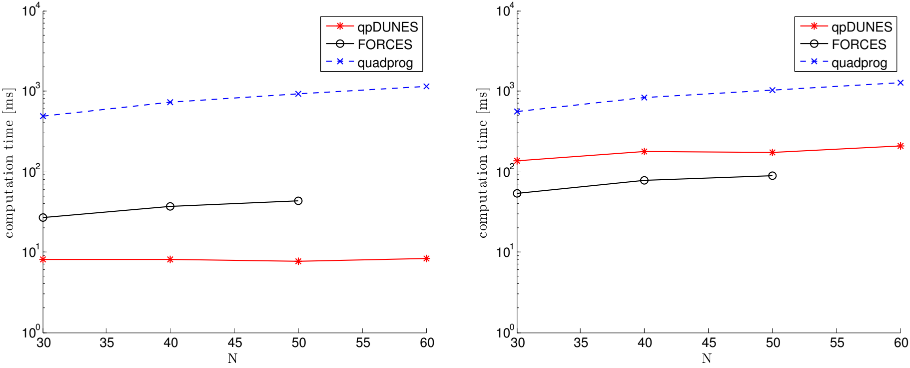
Double Integrator: Primal Regularization



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



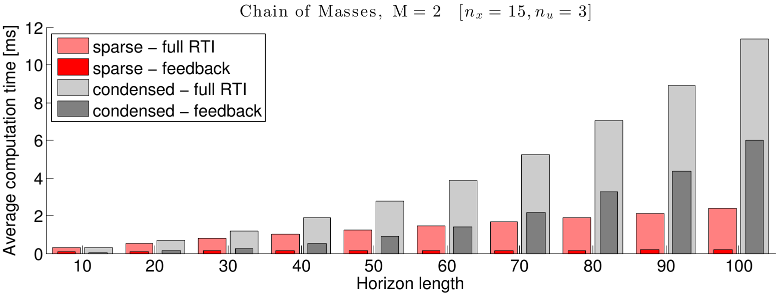
Hanging chain of masses: linear MPC $(M = 5)$

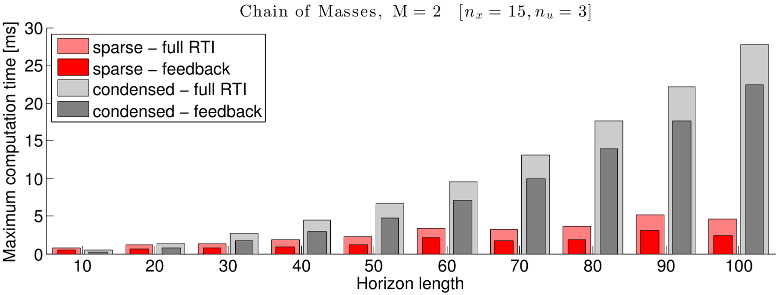


Mean computation times

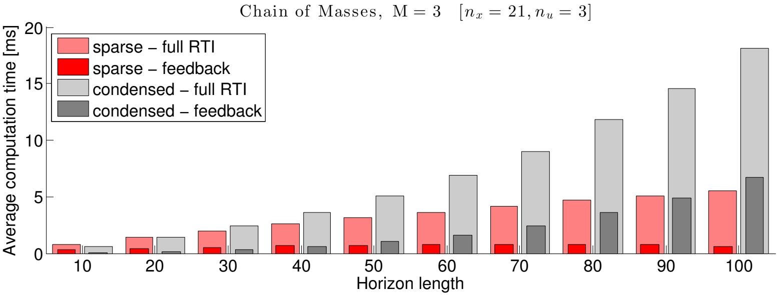
Maximum computation times

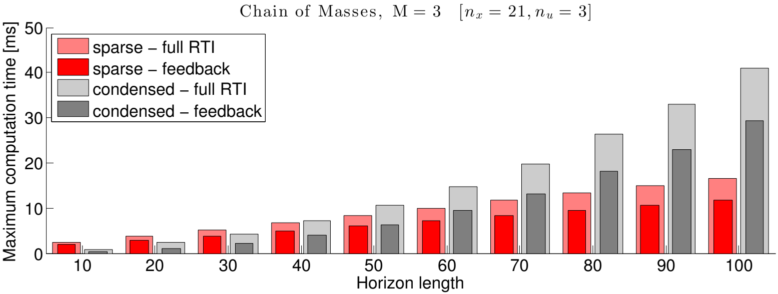
Hanging chain of masses: nonlinear MPC $(M = 2)$





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





qpDUNES roadmap

Current status

- linear MPC interfaces from C/C++ and Matlab
- available for nonlinear MPC in ACADO
- diagonal H_k, simple bounds: public; affine constraints: on request

Open theoretic isues

- infeasibility detection: only local proof and conjecture so far

Open software isues

- parallelization
- code generation & static memory version
- infeasibility detection

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