Iteration complexity analysis of dual first order methods: application to embedded and distributed MPC

Ion Necoara

Automatic Control and Systems Engineering Depart.
University Politehnica Bucharest
Outline

✓ Motivation
  • embedded MPC
  • distributed MPC
  • resource allocation in networks

✓ Dual first order algorithms
  • approximate primal solutions
  • convergence rate: suboptimality/infeasibility
  • numerical results

✓ Dual first order augmented Lagrangian algorithms
  • approximate primal solutions
  • convergence rate: suoptimality/infeasibility
  • numerical results

✓ Conclusions
Motivation I: embedded MPC

Embedded control requires:

• fast execution time ⇒ solution computed in very short time \((\sim ms)\)
• simple algorithm ⇒ suitable on cheap hardware ⇒ PLC, FPGA, ASIC, ...
• worst-case estimates for execution time for computing a solution ⇒ tight
• robust to low precision arithmetic ⇒ effects of round-off errors small

\[ \text{Embedded Model Predictive Control (MPC)} \]

• Linear systems: \(x_{t+1} = Ax_t + Bu_t\)
• State/input constraints: \(x_t \in X \& u_t \in U\) (\(X, U\) simple sets, e.g. box)
• Stage/final costs: \(\ell(x, u) \& \ell_f(x)\) (e.g. quadratic)
• Finite horizon optimal control of length \(N\):
  \[
  \min_{x_t \in X, u_t \in U} \sum_{t=0}^{N-1} \ell(x_t, u_t) + \ell_f(x_N) \\
  \text{s.t. : } x_{t+1} = Ax_t + Bu_t, \ x_0 = x
  \]
Optimization problem formulation

✓ Sparse formulation of MPC (i.e. without elimination of states):

\[
z = \begin{bmatrix} \mathbf{x}_1^T \cdots \mathbf{x}_N^T \mathbf{u}_0^T \cdots \mathbf{u}_{N-1}^T \end{bmatrix}^T \in \mathbb{R}^n \quad \& \quad Z = \prod_{t=1}^{N-1} \mathbb{X} \times \mathbb{X}_f \times \prod_{t=1}^N \mathbb{U}
\]

\[
f(z) = \sum_{t=0}^{N-1} \ell(x_t, u_t) + \ell_f(x_N)
\]

✓ MPC problem at state \(x\) formulated as primal convex problem with equality constraints:

\[
f^* = \min_{z \in \mathbb{R}^n} f(z)
\]

s.t.: \(Az = b, \quad z \in Z,\)

✓ Assumptions:

• \(f\) convex function (possibly nonsmooth & not strongly convex)
• \(Z\) simple convex set (e.g. box, \(\mathbb{R}^n\))
• \(Az = b\) equality constraints coming from dynamics
• difficult to project on the feasible set \(\{z \in Z : Az = b\}\)
Approaches for solving the convex problem

I. Primal methods

- interior-point/Newton methods [Rao’98], [Boyd’10], [Domahidi’12], [Kerrigan’10], [Patrinos’11], [N’09],...

- primal (sub)gradient/fast gradient methods [Richter’12], [Kogel’11],...

- active set methods [Ferreau’08], [Milman’08],...

- parametric optimization [Bemporad’02], [Tondel&Johansen’03], [Borelli’03], [Patrinos’10],...

II. Dual methods:

- dual (fast) gradient methods [Richter’11], [Patrinos’12], [M. Johansson’13], [N’08,12],...

- dual (fast) gradient augmented Lagrangian methods [Kogel’11], [N’12],...
Motivation II: distributed MPC

Distributed control requires:

- distributed computations \( \Rightarrow \) solution computed using only local information
- implementation on cheap hardware \( \Rightarrow \) simple schemes
- physical constraints on state/inputs \( \Rightarrow \) satisfied

\[ \Downarrow \]

Distributed Model Predictive Control (MPC)

- Coupling dynamics \( (M \) interconnected systems): 
  \[ x^{i}_{t+1} = \sum_{j \in \mathcal{N}^i} A^{ij}_x x^j_t + B^{ij}_u u^j_t \]
- Local state/input constraints: \( x^i_t \in X^i \) & \( u^i_t \in U^i \) 
  \((X^i, U^i \) simple sets)
- Local stage/final costs: \( \ell^i(x^i, u^i) \) & \( \ell^i_i(x^i) \)
- Finite horizon optimal control of length \( N \):
  \[ \min_{x^i_t \in X^i, u^i_t \in U^i} \sum_{i} \sum_{t} \ell^i(x^i_t, u^i_t) + \ell^i_i(x^i_N) \]
  \[ \text{s.t. : } x^i_{t+1} = \sum_{j \in \mathcal{N}^i} A^{ij}_x x^j_t + B^{ij}_u u^j_t, \ x_0 = x \]
Centralized optimization problem formulation

✓ Dense formulation of centralized MPC (i.e. elimination of states via dynamics):
✓ Define input trajectories for each subsystem:

\[
\mathbf{u}_i = [u^i_0 \cdots u^i_{N-1}] \quad \& \quad \mathbf{u} = [\mathbf{u}_1 \cdots \mathbf{u}_M]
\]

\[
f(\mathbf{u}) = \sum_i \sum_t \ell^i(x^i_t, u^i_t) + \ell^i_f(x^i_N)
\]

✓ Centralized MPC formulated as **primal** convex problem with inequality constraints:

\[
f^* = \min_{\mathbf{u}_i \in \mathbf{U}^i} f(\mathbf{u}_1, \cdots, \mathbf{u}_M) \iff \min_{\mathbf{u} \in \mathbf{U}} f(\mathbf{u})
\]

\[
s.t. : \quad g(\mathbf{u}_1, \cdots, \mathbf{u}_M) \leq 0 \quad \text{s.t. : } g(\mathbf{u}) \leq 0
\]

✓ Assumptions:

• function \( f \) **strongly** convex
• \( g \) convex coming from state constraints
• usually \( g(\cdot) \) is linear: \( g(\mathbf{u}) = G\mathbf{u} + g \) (separable in \( \mathbf{u}_i \) !)
• set \( \mathbf{U} = \mathbf{U}_1 \times \cdots \times \mathbf{U}_M \) convex & simple
• difficult to project on feasible set \( \{ \mathbf{u} \in \mathbf{U} : g(\mathbf{u}) \leq 0 \} \)
Motivation III: resource allocation

✓ Resource allocation problems in communication networks (e.g. Internet)

✓ Communication network

- set of traffic sources $S$
- set of links $\mathcal{L}$ with a finite capacity $c_l$
- each source associated with a route $r$ & transmit at rate $u_r$
- utility obtained by the source from transmitting data on route $r$ at rate $u_r$: $U_r(u_r)$

$$\begin{align*}
\max_{u_r \geq 0} & \sum_{r \in S} U_r(u_r) \\
\text{s.t.:} & \sum_{r:l \in r} u_r \leq c_l \quad \forall l \in \mathcal{L} \\
\downarrow & \\
\min_{u \in U} & f(u) \\
\text{s.t.:} & Gu + g \leq 0 \\
& g(u) \quad (g \text{ is a vector})
\end{align*}$$
Distributed approaches for solving the convex problem

I. Primal methods

- Jacobi type methods [Venkat’10], [Farina’12], [Scattolini’09], [Maestre’11], [Nesterov’10], [N’12],...
- penalty/interior point-methods [Camponogara’11,09], [Kozma’12],...
- gradient methods [Boyd’06], [N’13],...

II. Dual methods:

- dual Newton methods [Ozdaglar’10], [N’09,13],...
- dual gradient methods [Negenborn’08], [Doan’11], [Giselsson’12], [Rantzer’10], [Wakasa’08], [Foss’09], [N’08,12],...
- alternating direction methods [Boyd’11], [Conte’12], [Hansson’12], [Farokhi’12], [Koegel’12],...

\[\downarrow\]

usually dual methods cannot guarantee feasibility!
Brief history - first order methods

\[
\min_{u \in \mathbb{U}} f(u) \quad \& \quad \min_{u \in \mathbb{U}} f(u) \\
\text{s.t.} \quad A u = b \quad \text{s.t.} \quad g(u) \leq 0
\]

✓ First order methods: based on an oracle providing \( f(u) \) & \( \nabla f(u) \)

✓ “Simplest” first order method: Gradient Method

\[
\text{solutie ec. } F(u) = 0 \iff \text{fixed point iter. } u_{k+1} = u_k - \alpha F(u_k)
\]

• step size \( \alpha > 0 \) constant or variable
• simple iteration (vector operations)!
• fast/slow convergence?
• appropriate for \( x \) having very large dimension

• First derived by Cauchy (1847)
• Cauchy solved a nonlinear system of 6 equations with 6 unknowns

Brief history - first order methods

Slow rate of convergence for gradient method motivated work for finding other first order algorithms with faster convergence:

- **Conjugate Gradient Method** - independently proposed by Lanczos, Hestenes, Stiefel (1952)
  - convex QP: finds solution in $n$ iterations

- **Fast Gradient Method** - proposed by Yurii Nesterov (1983)
  - one order faster than classic gradient

- FGM unused for 2 decades! - now one of the most used optimization methods in small/large applications

- Google returns approximately 20 mil. results ($\approx 2000$ citations)
Gradient method (GM)

\[
\min_{u \in \mathbb{R}^n} f(u)
\]

- Gradient method (GM) for optimization problem:
  \[
  u_{k+1} = u_k - \alpha_k \nabla f(u_k)
  \]

- Step size \(\alpha_k\) can be chosen as: constant, Wolfe conditions, backtracking, ideal, ...

- Advantages of GM:
  - reduced complexity per iteration - \(O(n)\) + computation of \(\nabla f(u)\)
  - does not use Hessian information
  - global convergence under usual conditions
  - robust to errors from computations/inexact gradients [Dev:13],[Nec:14]

- Disadvantages of GM:
  - slow convergence - sublinear/at most linear (under regularity conditions)


Rate of convergence for GM

Assume \( f \) is convex and gradient \( \nabla f(u) \) Lipschitz, i.e.

\[
\| \nabla f(u) - \nabla f(v) \| \leq L \| u - v \| \quad \forall u, v \in \text{dom} f
\]

Gradient method (MG) with constant step size \( \alpha = 1/L \)

\[
u_{k+1} = u_k - 1/L \nabla f(u_k)
\]

**Theorem.** Under convexity and Lipschitz gradient, GM has sublinear convergence:

\[
f(u_k) - f^* \leq \frac{L\|u_0 - u^*\|^2}{2k}
\]

**Theorem.** If additionally \( f \) is strongly convex with constant \( \sigma \), i.e.

\[
f(v) \geq f(u) + \langle \nabla f(u), v - u \rangle + \frac{\sigma}{2} \| u - v \|^2 \quad \forall u, v
\]

then GM has linear rate of convergence:

\[
f(u_k) - f^* \leq \left( \frac{L - \sigma}{L + \sigma} \right)^k \frac{L\|u_0 - u^*\|^2}{2}
\]
Fast gradient method (FGM)

Slow convergence of GM $\implies$ develop new methods with better performance:

- Fast Gradient Method (Nesterov 1983) - one order faster than classical GM

**Fast Gradient Method iteration:**

- Gradient step: $u_{k+1} = v_k - \frac{1}{L} \nabla f(v_k)$
- Linear combination: $v_{k+1} = u_{k+1} + \theta_k (u_{k+1} - u_k)$

- Initial points $u_0 = v_0$

- $\theta_k$ chosen appropriately, e.g. for $f$ strong convex $\theta_k = \frac{\sqrt{L} - \sqrt{\sigma}}{\sqrt{L} + \sqrt{\sigma}}$.

- FGM has better performance than GM, but complexity per iteration remains comparable with GM and thus low (FGM is optimal!).

## Convergence GM versus FGM

<table>
<thead>
<tr>
<th>Conditions $f$</th>
<th>GM</th>
<th>FGM</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f(u)$ convex &amp; $\nabla f(u)$ Lipschitz</td>
<td>$O\left(\frac{LR^2}{k}\right)$</td>
<td>$O\left(\frac{LR^2}{k^2}\right)$</td>
</tr>
<tr>
<td>$\nabla f(u)$ Lipschitz &amp; $f(u)$ strong convex</td>
<td>$O\left(\left(\frac{L-\sigma}{L+\sigma}\right)^k\right)$</td>
<td>$O\left(\left(1 - \sqrt{\frac{\sigma}{L}}\right)^k\right)$</td>
</tr>
</tbody>
</table>

![Graph showing convergence](image)

**Legend:**
- **MG(L)**
- **MGA(L)**
- **MG(L,\sigma)**
- **MGA(L,\sigma)**
Gradient type methods: primal vrs. dual

• Primal optimization problem:
\[
\min_{u \in U} f(u)
\]

• Iterations of first order methods need to remain feasible $\rightarrow$ projected gradient
\[
{u}_{k+1} = [{u}_k - \alpha_k \nabla f(u_k)]_U
\]

• Major advantage primal approach: under $\nabla f(u)$ Lipschitz $\Rightarrow$ sublinear convergence; additionally strong convexity $\Rightarrow$ linear convergence for projected GM

• Major disadvantage primal approach: need to project on $U$: $[u_k - \alpha_k \nabla f(u_k)]_U$

• If $U$ not simple set (e.g. polyhedron described by $Gu + g \leq 0 \& Au = b$, with $G$ and $A$ general matrices), then projection is difficult $\rightarrow$ dual approach
Gradient type methods: primal vrs. dual

• Dual approach: solve dual problem

\[
\max_{x \in X} \ d(x) \quad \text{cu} \quad X = \mathbb{R}^n_+
\]

• Major advantage against primal approach: projection need to be performed only for Lagrange multipliers corresponding to inequalities \( x \in \mathbb{R}^n_+ \) - simple!

• Major disadvantage of dual gradient type methods:
  I. how to recover approximate primal solution [Nec:13], [Nec:14]
  II. iteration complexity of dual gradient type methods: sublinear
      \( \mathcal{O} \left( \frac{1}{k^p} \right), \ p = 1, 2 \) [Nec:13], [Nec:14] or (local) linear convergence ([LuoTse:93]) [Nec:14]

I and II - main motivations for the work presented here!


PART I
Inequality constrained problems - dual formulation

\[ f^* = \min_{u \in U} \{f(u) : g(u) \leq 0\} \]

Primal convex problem, where:

- \( f \) strong convex (e.g. \( \nabla^2 f \succeq \sigma I_n \))
- \( g \) convex & \( \|\nabla g\|_F \leq c_g \) (e.g. \( g(u) = Gu + g \))
- \( U \) simple set (e.g. box, ball,…) & strong duality holds

**GOAL** - for desired accuracy \( \epsilon_{out} \) compute nr. iterations \( k \) and generate \( \hat{u}^k \in U \):

\[ |f(\hat{u}^k) - f^*| \leq O(\epsilon_{out}) \quad \& \quad \|[g(\hat{u}^k)]_+\| \leq O(\epsilon_{out}) \]

**Approach** - use the dual formulation

\[ d(x) = \min_{u \in U} \mathcal{L}(u, x) \quad (:= f(u) + \langle x, g(u) \rangle) \]

- Define \( u(x) \) solution of inner problem: \( u(x) \in \arg \min_{u \in U} \mathcal{L}(u, x) \), with \( U \) - SIMPLE!

- **Outer probl.** (smooth): \( \max_{x \in \mathbb{R}^m_+} d(x) \Rightarrow \) solve outer problem with dual first order met.!
Dual formulation: properties

\[ f^* = \min_{u \in U} \{ f(u) : g(u) \leq 0 \} \iff \max_{x \in \mathbb{R}^m_+} d(x) \]

where \( u(x) \) solution of inner problem: \( u(x) \in \arg \min_{u \in U} L(u, x) \), with \( U - SIMPLE! \)

\[ d(x) = \min_{u \in U} L(u, x) \quad (:= f(u) + \langle x, g(u) \rangle) \]

**Lemma 1** Dual function \( d(x) \) satisfies:

- if \( f \) strongly convex \( (\sigma_f > 0) \) \( \Rightarrow \) the dual \( d(x) \) differentiable: \( \nabla d(x) = g(u(x)) \)
- if \( \| \nabla g(u) \|_F \leq c_g \) for all \( u \in U \) \( \Rightarrow d(x) \) has Lipschitz gradient
- Lipschitz constant \( L_d = \frac{c_g^2}{\sigma_f} \) (or \( L_d = \frac{\| G \|^2}{\sigma_f} \) for \( g(u) = Gu + g \))

**Lemma 2** Dual function \( d(x) \) and inner solution \( u(x) \) satisfy:

- primal suboptimality 1: \( \frac{\sigma_f}{2} \| u(x) - u^* \|^2 \leq f^* - d(x) \quad \forall x \geq 0 \)
- primal suboptimality 2: \( |f(u(x)) - f^*| \leq c_g (\| x - x^* \| + \| x^* \|) \| u(x) - u^* \| \)
- primal feasibility: \( \|[g(u(x))]_+\| \leq c_g \| u(x) - u^* \| \)
Dual first order methods

\[ f^* = \min_{u \in U} \{ f(u) : g(u) \leq 0 \} \iff \max_{x \in \mathbb{R}_+^m} d(x) \]

**Lemma 1 state:** if \( f \) strong convex and \( \| \nabla g \|_F \leq c_g \), then \( \nabla d \) Lipschitz i.e. \( \max_{x \in \mathbb{R}_+^m} d(x) \) smooth \( \Rightarrow \) descent lemma valid:

\[ d(x) + \langle \nabla d(x), y - x \rangle - \frac{L_d}{2} \| y - x \|^2 \leq d(y) \quad \forall x, y \in \mathbb{R}_+^m \]

Dual first order alg.: updates two dual sequences \((x^k, y^k)\) and one primal sequence \(u^k\):

**Algorithm (DFO)**

Given \( x^0 = y^1 \in X \), for \( k \geq 1 \) compute:

1. \( u^k = \arg \min_{u \in U} \mathcal{L}(u, y^k) \)
2. \( x^k = [y^k + \alpha_k \nabla d(y^k)]_+ \)
3. \( y^{k+1} = x^k + \frac{\theta_k - 1}{\theta_{k+1}} (x^k - x^{k-1}) \)

- **Dual Gradient (DG):** parameters (step size) \( \frac{1}{L_d} \leq \alpha_k \leq \frac{1}{L_D} \) and \( \theta_k = 1 \)
- **Dual Fast Gradient (DFG):** parameters \( \alpha_k = \frac{1}{L_d} \) and \( \theta_{k+1} = \frac{1 + \sqrt{1 + 4 \theta_k^2}}{2} \)
Dual gradient alg. (DG)

**Dual Gradient (DG):** chooses parameters \( \frac{1}{L_d} \leq \alpha_k \leq \frac{1}{L_D} \) and \( \theta_k = 1 \) in Alg. (DFO)

\[
\begin{align*}
(DG): & \quad x^{k+1} = \left[ x^k + \alpha_k \nabla d(x^k) \right]_+
\end{align*}
\]

- dual gradient: \( \nabla d(x^k) = g(u^k) \)
- \( u^k = u(x^k) \) exact solution of the inner problem for given \( x^k \), i.e.
  \[
  u^k = \arg \min_{u \in U} L(u, x^k)
  \]

✓ Define dual and primal last iterate sequences:

\[
(x^k \quad \& \quad u^k)
\]

✓ Define dual and primal average sequences:

\[
\begin{align*}
(x^k \quad \& \quad \hat{u}^k = \frac{\sum_{j=0}^{k} \alpha_j u^j}{S^k} \quad \text{with} \quad S^k = \sum_{j=0}^{k} \alpha_j)
\end{align*}
\]
Convergence rate - estimates for suboptimality/infeas.

Estimates for suboptimality and feasibility violation for approximate primal and dual solutions in average \((\hat{u}^k, \hat{x}^k)\) and last iterate \((u^k, x^k)\) generated by \((\text{DG})\)

- Dual suboptimality: \(f^* - d(x^k) \leq O\left(\frac{L_d R_d^2}{k}\right)\)

**Theorem 1**: for approximate primal solution in average \(\hat{u}^k\)

- Primal suboptimality: \(|f(\hat{u}^k) - f^*| \leq O\left(\frac{L_d R_d^2}{k}\right)\)
- Primal feasibility violation: \(\|[h(\hat{u}^k)]_+\| \leq O\left(\frac{L_d R_d}{k}\right)\)

**Theorem 2**: for approximate primal solution in last iterate \(u^k\)

- Primal suboptimality: \(|f(u^k) - f^*| \leq O\left(\frac{L_d R_d^2}{\sqrt{k}}\right)\)
- Primal feasibility violation: \(\|[h(u^k)]_+\| \leq O\left(\frac{L_d R_d}{\sqrt{k}}\right)\)

Here \(R_d = \min_{x^* \in X^*} \|x^0 - x^*\|\).
Dual fast gradient alg. (DFG)

Dual Fast Gradient (DFG): choose \( \alpha_k = \frac{1}{L_d} \) and \( \theta_{k+1} = \frac{1+\sqrt{1+4\theta_k^2}}{2} \) in Alg. (DFO)

\[
\begin{align*}
\text{DFG} & \quad \left\{ \begin{array}{l}
x^k = \left[ y^k + \frac{1}{L_d} \nabla d(y^k) \right]_+ \\
y^{k+1} = x^k + \frac{\theta_{k-1}}{\theta_{k+1}} (x^k - x^{k-1})
\end{array} \right.
\end{align*}
\]

- dual gradient: \( \nabla d(y^k) = g(u^k) \)
- \( u^k = u(y^k) \) exact solution of inner problem: \( u^k = \arg \min_{u \in U} \mathcal{L}(u, y^k) \)

✓ Define dual and primal last iterate sequences:

\[
\begin{align*}
(x^k \quad \& \quad v^k = u(x^k) = \arg \min_{u \in U} \mathcal{L}(u, x^k))
\end{align*}
\]

✓ Define dual and primal average sequences:

\[
\begin{align*}
(x^k \quad \& \quad \hat{u}^k = \frac{\sum_{j=0}^{k} \theta_j u^j}{S^k_{\theta}} \quad \text{with} \quad S^k_{\theta} = \sum_{j=0}^{k} \theta_j)
\end{align*}
\]
Convergence rate - estimates for suboptimality/infeas.

Estimates for suboptimality and feasibility violation for approximate primal and dual solutions in average \((\hat{u}^k, x^k)\) and last iterate \((v^k, x^k)\) generated by \((DFG)\)

- \textbf{Dual suboptimality:} \[ f^* - d(x^k) \leq O\left( \frac{L_d R_d^2}{k^2} \right) \]

**Theorem 3:** for approximate primal solution in average \(\hat{u}^k\)

- \textbf{Primal suboptimality:} \[ |f(\hat{u}^k) - f^*| \leq O\left( \frac{L_d R_d^2}{k^2} \right) \]
- \textbf{Primal feasibility violation:} \[ \| [g(\hat{u}^k)]^+ \| \leq O\left( \frac{L_d R_d}{k} \right) \]

**Theorem 4:** for approximate primal solution in last iterate \(v^k\)

- \textbf{Primal suboptimality:} \[ |f(v^k) - f^*| \leq O\left( \frac{L_d R_d^2}{k} \right) \]
- \textbf{Primal feasibility violation:} \[ \| [g(v^k)]^+ \| \leq O\left( \frac{L_d R_d}{k} \right) \]

Similarly, \[ R_d = \min_{x^* \in X^*} \|x^0 - x^*\| \].
Summary - estimates for suboptimality/infeas. Alg. (DFO)

✓ Estimates on suboptimality & infeasibility depend on 3 constants:

- initial starting point \( x^0 \)
- its distance to the optimal dual set \( X^* : R_d \)
- Lipschitz constant of the dual: \( L_d = \frac{c^2}{\sigma_f} \)

✓ Summary of estimates of primal suboptimality & infeasibility for (DFO):

<table>
<thead>
<tr>
<th>Alg.</th>
<th>DG</th>
<th>DFG</th>
</tr>
</thead>
<tbody>
<tr>
<td>last</td>
<td>( O \left( \frac{1}{\sqrt{k}} \right) )</td>
<td>( O \left( \frac{1}{k} \right) )</td>
</tr>
<tr>
<td>average</td>
<td>( O \left( \frac{1}{k} \right) )</td>
<td>( O \left( \frac{1}{k^2} \right) )</td>
</tr>
</tbody>
</table>

✓ From our practical experience, usually (DFO) algorithms perform better in the primal last iterate that in an average of iterates (when the problem is well conditioned)!
Numerical results I

\[ \min_{u \geq 0, \, Gu + g \leq 0} \ u^T Q u + q^T u + \gamma \log(1 + a^T u), \quad G \in \mathbb{R}^{1.5n \times n}, \, \sigma = 1, \, \gamma = 0.1 \]

- typical behavior of (DFO) methods: oscillating in primal feasibility & suboptimality!
- dual first order methods performs better in last iterates than in average sequences (worst case analysis says differently)!
Numerical results II

\[
\min_{u>0, G'u+q<0} \quad u^T Qu + q^T u + \gamma \log(1 + a^T u), \quad G \in \mathbb{R}^{1.5n \times n}, \sigma = 1, \gamma = 0.1
\]

- number of iterations of dual first order methods for 30 random test cases: fix dimension & variable dimension
- number of iterations not varying much for different test cases of fix dimension
- number of iterations are mildly dependent in problem dimension
Numerical results III

\[
\min_{lb \leq u \leq ub, \ G u + g \leq 0} \ u^T Q u + q^T u + \gamma \log(1 + \exp^{\sigma^T u}), \quad G \in \mathbb{R}^{1.5n \times n}, \ \sigma = 1, \ \gamma = 0.1
\]

<table>
<thead>
<tr>
<th>Alg./n</th>
<th>10</th>
<th>50</th>
<th>$10^2$</th>
<th>$10^3$</th>
<th>$5 \times 10^3$</th>
<th>$10^4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_{DG}$ last</td>
<td>35</td>
<td>195</td>
<td>463</td>
<td>782</td>
<td>1.147</td>
<td>2.155</td>
</tr>
<tr>
<td>$k_{DG}$ avg.</td>
<td>527</td>
<td>3.423</td>
<td>12.697</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$k_{DFG}$ last</td>
<td>19</td>
<td>61</td>
<td>97</td>
<td>198</td>
<td>276</td>
<td>292</td>
</tr>
<tr>
<td>$k_{DFG}$ avg.</td>
<td>41</td>
<td>108</td>
<td>186</td>
<td>381</td>
<td>563</td>
<td>582</td>
</tr>
</tbody>
</table>

- average results for 10 random problems ($\epsilon = 10^{-2}$)
- $Q$ and $G$ sparse for large problems (few tens $\approx 50$ of nonzeros on each row)
- dual first order methods performs better in last iterates than in average sequences (worst case analysis says differently)!
Numerical results IV

\[
\begin{align*}
\min_{u \geq 0, \ G u + g \leq 0} & \quad u^T Q u + q^T u + \gamma \log(1 + a^T u), \\
G & \in \mathbb{R}^{1.5n \times n}, \ \sigma = 1, \ \gamma = 0.1
\end{align*}
\]

- practical performance usually better in the last iterate
- however, there are also cases where the practical performance is comparable with the theoretical estimates
Numerical results V

→ our estimates are dependent on $x^*$ via $R_d = \| x^0 - x^* \|\ldots$ BUT no good bounds on $x^*$ (e.g. bounds based on a Slater vector [Nedich'09] or [Patrinos:14])!

→ average number of outer iterations (theoretical and real) on 10 random generated QPs for accuracy $\epsilon = 10^{-3}$ and variable dimension $n$

→ plots show the behaviour of Alg. (DG) and (DFG) in average of primal iterates
Approximate dual function & gradient

\[ \max_{x \in \mathbb{R}^m_+} d(x) \Rightarrow \text{solved with (DFO)} \Rightarrow \text{convergence rate } O \left( \frac{1}{\sqrt{k}} \right) \text{ to } O \left( \frac{1}{k^2} \right) \]

- BUT... gradient \( \nabla d(x) \) requires EXACT solution of inner \( \Rightarrow \) hard to compute
- Possible remedy: approximate solution of inner problem \( \Rightarrow \) inexact dual gradient

\[ \tilde{u}(x) \approx \arg \min_{u \in U} f(u) + \langle x, g(u) \rangle \]

such that the following stopping criterion holds

\[ \tilde{u}(x) \in U, \quad \mathcal{L}(\tilde{u}(x), x) - \mathcal{L}(u(x), x) \leq \epsilon_{\text{in}}/3 \]

- Introduce two notions: inexact dual function and gradient

\[ \bar{d}(x) = \mathcal{L}(\tilde{u}(x), x) \quad \& \quad \nabla \bar{d}(x) = g(\tilde{u}(x)) \]

**Lemma 3** (extension of result in [Devolder'11] from linear \( g \) to general convex \( g \)):
Based on the above stopping criterion, we have for all \( x, y \in \mathbb{R}^m_+ \):

\[ \bar{d}(x) + \langle \nabla \bar{d}(x), y - x \rangle - L_d \| y - x \|^2 - \epsilon_{\text{in}} \leq d(y) \leq \bar{d}(x) + \langle \nabla \bar{d}(x), y - x \rangle \]
Convergence rate of (DFO) under inexact dual gradients

Estimates on suboptimality and infeasibility for approximate primal and dual solutions in average \((\hat{u}^k, x^k)\) generated by INEXACT (DFO) Alg.: \(\nabla \bar{d}(x^k) = g(\bar{u}^k)\)

**Theorem 5:** approximate primal and dual solutions in average \((\hat{u}^k, x^k)\) of inexact (DG)

- Dual suboptimality: \(f^* - d(x^k) \leq O\left(\frac{L_d R_d^2}{k}\right) + \epsilon_{in}\)
- Primal suboptimality: \(|f(\hat{u}^k) - f^*| \leq O\left(\frac{L_d R_d^2}{k}\right) + \epsilon_{in}\)
- Primal feasibility violation: \(\|[g(\hat{u}^k)]_{+}\| \leq O\left(\frac{L_d R_d}{k}\right) + O\left(\sqrt{\frac{\epsilon_{in}}{k}}\right)\)

**Theorem 6:** approximate primal and dual solutions in average \((\hat{u}^k, x^k)\) of inexact (DFG)

- Dual suboptimality: \(f^* - d(x^k) \leq O\left(\frac{L_d R_d^2}{k^2}\right) + k\epsilon_{in}\)
- Primal suboptimality: \(|f(\hat{z}^k) - f^*| \leq O\left(\frac{L_d R_d^2}{k^2}\right) + k\epsilon_{in}\)
- Primal feasibility violation: \(\|[h(\hat{u}^k)]_{+}\| \leq O\left(\frac{L_d R_d}{k^2}\right) + O\left(\sqrt{\frac{\epsilon_{in}}{k}}\right)\)
Convergence rate - estimates for suboptimality/infeas.

For a desired outer accuracy $\epsilon_{out}$ we can choose in inexact (DFO) alg.:

<table>
<thead>
<tr>
<th>Alg.</th>
<th>inexact (DG)</th>
<th>( k_{out} = \mathcal{O} \left( \left\lfloor \frac{L_d R_d^2}{\epsilon_{out}} \right\rfloor \right) ) &amp; $\epsilon_{in} = \epsilon_{out}$</th>
<th>inexact (DFG)</th>
<th>( k_{out} = \mathcal{O} \left( \left\lfloor R_d \sqrt{\frac{L_d}{\epsilon_{out}}} \right\rfloor \right) ) &amp; $\epsilon_{in} = \epsilon_{out} \sqrt{\epsilon_{out}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>average</td>
<td>[ f(\hat{u}<em>k^{k</em>{out}}) - f^* \leq \epsilon_{out} ] &amp; [ \epsilon_{in} = \epsilon_{out} ]</td>
<td>[ | g(\hat{u}<em>k^{k</em>{out}}) | + \leq \epsilon_{out} ]</td>
<td>(DFG) more sensitive to errors (inexact gradients) than (DG)</td>
<td></td>
</tr>
</tbody>
</table>
Numerical results VI

- nr. outer iterations on random QP for $\epsilon_{\text{out}} = 10^{-3}$ & inner accuracy $\epsilon_{\text{in}}$ variable
- (DFG) better than (DG) in average primal sequence ($O(\frac{1}{k^2})$ instead of $O(\frac{1}{k})$)
- BUT inner problem has to be solved with higher accuracy in (DFG) than in (DG)
Issues on strong convexity for dual function $d$

We consider smooth (i.e. Lipschitz gradient objective function) convex problem:

$$(P): \min_{\mathbf{u}} f(\mathbf{u}) \quad \text{s.t.} \quad g(\mathbf{u}) \leq 0$$

- When we have linear convergence of first order methods for above primal problem? ⇒ Answer: e.g. when $f$ is strong convex & gradient Lipschitz
- When we have linear convergence of dual first order methods for the dual of primal problem (P)? ⇒ Answer: error bound type property
- In practice: strong convexity for $f$ is not found often in applications

**Example 1:**

$$f(x) = x^T Q x \Rightarrow L = \|Q\| = \lambda_{\max}(Q) \quad \& \quad \sigma = 1/\|Q^{-1}\| = \lambda_{\min}(Q)$$

**Example 2:**

$$f(x) = \log(1 + e^{a^T x}) \Rightarrow L = \|\nabla^2 f(x)\| = \left\| \frac{e^{a^T x}}{(1 + e^{a^T x})^2} a a^T \right\| \leq \frac{\|a\|^2}{4} \quad \& \quad \sigma = 0$$
Issues on strong convexity for dual function \( d \)

We consider smooth (i.e. Lipschitz gradient objective function) convex problem:

\[
\min_{u: \, Gu + g \leq 0} f(u) \quad \& \quad U = \mathbb{R}^n
\]

**Classic**: for proving linear convergence in primal first order methods it is sufficient that \( f \) is strong convex & Lipschitz gradient

- \( f \) strong convex, then dual fct. \( d(x) = \min_{u \in \mathbb{R}^n} f(u) + \langle x, Gu + g \rangle \) Lipschitz gradient
- If \( f \) Lipschitz gradient, then dual function \( d \) is not strongly convex!
- Therefore, primal gradient converges linearly on primal problem, while dual gradient converges sublinearly on dual problem!

**gap in the primal-dual methods!**

However, many applications modelled as \( \min_{x \in X} f(x) \), where:

- objective function \( f \) in the form \( f(x) = g(Ax) \), with \( g(\cdot) \) strong convex
- constraints set \( X \) polyhedral: \( X = \{ x : \, Cx \leq c \} \)
- **Example**: \( d(x) = \min_{u \in \mathbb{R}^n} f(u) + \langle x, Gu \rangle = -\tilde{f}(-G^T x) \). Here \( \tilde{f} \) is conjugate function of \( f \) and is strong convex provided that \( f \) has Lipschitz gradient!
Linear convergence of dual gradient method

\[ (P) : \min_{u \in \mathbb{R}^n} \{ f(u) \, : \, G u + g \leq 0 \} \]

Assume \( f \) is strong convex & Lipschitz gradient

- Primal first order methods converge linearly!
- \( f \) strong convex, then dual fct. \( d(x) = \min_{u \in \mathbb{R}^n} \{ f(u) + \langle x, G u + g \rangle \} \) Lipschitz gradient
- If \( f \) Lipschitz gradient, then dual function \( d \) is not strongly convex! BUT \( d(x) = -\tilde{f}(-G^T x) + g^T x \). Here \( \tilde{f} \) is conjugate function of \( f \) and is strong convex provided that \( f \) has Lipschitz gradient

In [Nec:14] we proved that problems of the form: \( \min_{x \in X} f(x) \), where:

- \( f \) in the form \( f(x) = g(Ax) + b^T x \), with \( g(\cdot) \) strong convex & Lip. gradient
- constraints set \( X \) polyhedral: \( X = \{ x : C x \leq c \} \)
- Note: corresponding dual problem of (P) has these assumptions!

**SATISFY AN ERROR BOUND PROPERTY**

error bound property allows to prove linear convergence of dual gradient method!

Numerical results VII

\[ \min_{u \in \mathbb{R}^n : Gu + g \leq 0} u^T Q u + q^T u + \gamma \log(1 + \exp^{a^T u}), \quad G \in \mathbb{R}^{1.5n \times n}, \sigma = 1, \gamma = 0.1 \]

- linear convergence of (DG) in last iterate for \( U = \mathbb{R}^n \): logarithmic scale of primal suboptimality and infeasibility
- compare with the theoretical sublinear estimates (dot lines) for the convergence rate \( O\left(\frac{1}{\sqrt{k}}\right) \)
- plot clearly shows our theoretical findings, i.e. linear convergence
MPC - feasibility & stability

Feasibility: guaranteed by combining our theory for suboptimality/feasibility violation with constraint tightening

\[ f^* = \min_{{u \in U}} \{ f(u) : g(u) + \epsilon_c e \leq 0 \} \]

**GOAL** - for desired accuracy \( \epsilon_{\text{out}} \) compute \( k_{\text{out}}, \epsilon_{\text{in}} \) and \( \epsilon_c \) and generate \( \hat{u}^{k_{\text{out}}} \in U \):

\[ |f(\hat{u}^{k_{\text{out}}}) - f^*| \leq O(\epsilon_{\text{out}}) \quad \& \quad g(\hat{u}^{k_{\text{out}}}) \leq 0 \]

\[ \Downarrow \]

Slater vector approach

- inexact (DG) average primal sequence
- \( k_{\text{out}} = O\left(\frac{1}{\epsilon_{\text{out}}}\right) \), \( \epsilon_{\text{in}} \approx \epsilon_{\text{out}} \) and \( \epsilon_c \approx \epsilon_{\text{out}} \)

- inexact (DFG) average primal sequence
- \( k_{\text{out}} = O\left(\sqrt{\frac{1}{\epsilon_{\text{out}}}}\right) \), \( \epsilon_{\text{in}} \approx \epsilon_{\text{out}}\sqrt{\epsilon_{\text{out}}} \) and \( \epsilon_c \approx \epsilon_{\text{out}} \)

Stability: follows from stability of suboptimal MPC with quadratic cost by choosing \( \epsilon_{\text{out}} \) adequately:

\[ \epsilon_{\text{out}}^+ \leq \min \left\{ \|x\|^2_Q, c \cdot \min_j \{ g_j(x^+, \tilde{u}^+) \} \right\} \]

where \( x^+ \) is the next state in the MPC scheme and \( \tilde{u}^+ \) is a corresponding Slater vector
PART II
Equality constrained problems: dual formulation

\[ f^* = \min_{z \in Z} \{ f(z) : Az = b \} \]

✓ Primal convex problem, where:

- \( f \) only convex (e.g. \( \nabla^2 f \succeq 0 \))
- \( Z \) simple set (e.g. box, ball,...) & strong duality holds

**GOAL** - for desired accuracy \( \epsilon_{out} \) compute nr. iterations \( k \) and generate \( \hat{z}^k \in Z \):

\[ |f(\hat{z}^k) - f^*| \leq \mathcal{O}(\epsilon_{out}) \quad \& \quad \|A\hat{z}^k - b\| \leq \mathcal{O}(\epsilon_{out}) \]

✓ Dual function \( d(x) = \min_{z \in Z} L(z, x) \) (\( := f(z) + \langle x, Az - b \rangle \))

- \( f \) strictly convex \( \Rightarrow d(x) \) differentiable
- \( f \) just convex \( \Rightarrow d(x) \) nonsmooth: \( Az(x) - b \in \partial d(x) \)

where \( z(x) \) solution of **inner problem**: \( z(x) \in \arg \min_{z \in Z} L(z, x) \), with \( Z \) - SIMPLE!

✓ **Outer problem** \( \max_{x \in \mathbb{R}^m} d(x) \) nonsmooth \( \Rightarrow \) solve outer with subgradient alg. \( \Rightarrow \) slow convergence \( \mathcal{O} \left( \frac{1}{\sqrt{k}} \right) \)

**Approach** - *Augmented Lagrangian* (method of multipliers) by Hestenes/Powell'69
Augmented dual function

✓ Augmented dual function:

\[
d_\rho(x) = \min_{z \in Z} \mathcal{L}_\rho(z, x) \quad (:= f(z) + \langle x, Az - b \rangle + \frac{\rho}{2} \|Az - b\|^2)
\]

✓ Function \(d_\rho\) satisfies:

• concave
• differentiable with gradient \(\nabla d_\rho(x) = Az(x) - b\)
• inner problem \(z(x) = \arg \min_{z \in Z} [f(z) + \langle x, Az - b \rangle + \frac{\rho}{2} \|Az - b\|^2]\)
• gradient \(\nabla d_\rho\) Lipschitz with \(L_d = \frac{1}{\rho}\)

✓ Outer problem \(\max_{x \in \mathbb{R}^m} d_\rho(x)\) smooth \(\Rightarrow\) solved with dual first order methods (DFO) \(\Rightarrow\) convergence rate of order \(O\left(\frac{1}{k}\right)\) or \(O\left(\frac{1}{k^2}\right)\)

• BUT gradient \(\nabla d_\rho(x)\) requires EXACT solution of inner problem \(\Rightarrow\) hard to compute
• Possible remedy: approximate solution of inner problem \(\Rightarrow\) inexact dual gradient
Inexact augmented dual function & gradient

\[ \bar{z}(x) \approx \arg \min_{z \in Z} f(z) + \langle x, Az - b \rangle + \frac{\rho}{2} \|Az - b\|^2 \]

such that one of the following stopping criterions hold

\[ \bar{z}(x) \in Z, \quad \mathcal{L}_\rho(\bar{z}(x), x) - \mathcal{L}_\rho(z(x), x) \leq O(\epsilon_{in}^2) \]

or

\[ \bar{z}(x) \in Z, \quad \langle \nabla \mathcal{L}_\rho(\bar{z}(x), x), z - \bar{z}(x) \rangle \geq -\epsilon_{in} \forall z \in Z \]

Define two notions:

- inexact dual function value: \( \bar{d}_\rho(x) = \mathcal{L}_\rho(\bar{z}(x), x) \)
- inexact dual gradient: \( \nabla \bar{d}_\rho(x) = A\bar{z}(x) - b \)

**Lemma**: \( \|\nabla \bar{d}_\rho(x) - \nabla d_\rho(x)\| \leq O(\epsilon_{in}) \)

**Lemma** [Devolder’11]: Based on one stopping criterion from above, we have \( \forall x, y \in \mathbb{R}^m \)

\[ \bar{d}_\rho(x) + \langle \nabla \bar{d}_\rho(x), y - x \rangle - \frac{L_d}{2} \|y - x\|^2 - \epsilon_{in} \leq d_\rho(y) \leq \bar{d}_\rho(x) + \langle \nabla \bar{d}_\rho(x), y - x \rangle \]
Inexact dual gradient augm. Lagrangian alg.

From previous descent type lemma ⇒ smooth dual ⇒ dual first order methods

$$ (\text{IDGAL}) : \quad x^{k+1} = x^k + \alpha^k \nabla \tilde{d}_\rho (x^k) $$

where

- $\nabla \tilde{d}_\rho (x^k) = A\tilde{z}^k - b$ inexact dual gradient
- $\tilde{z}^k = \tilde{z}(x^k)$ approximate solution of the inner problem for given $x^k$
  $$ \tilde{z}^k \approx \arg \min_{z \in Z} f(z) + \langle x^k, Az - b \rangle + \frac{\rho}{2} \| Az - b \|^2 $$
- REMARK: the theory works also for $\epsilon_{in} = 0$ (i.e. inner problem solved exactly) or for $Z = \mathbb{R}^n$ (i.e. inner problem is unconstrained)!

Define the average dual and primal sequences $(\hat{x}^k, \hat{z}^k)$:

$$ \hat{x}^k = \frac{\sum_{j=0}^{k} \alpha^j x^{j+1}}{S^k} \quad \& \quad \hat{z}^k = \frac{\sum_{j=0}^{k} \alpha^j \tilde{z}^j}{S^k}, \quad \text{with} \quad S^k = \sum_{j=0}^{k} \alpha^j $$
**Convergence rate - estimates for suboptimality/infeas.**

**Main result:** estimates for suboptimality and feasibility violation for the approximate primal and dual solutions in average $\hat{z}^k$ and $\hat{x}^k$ generated by (IDGAL)

**Theorem 7:**

- Dual suboptimality: $f^* - d_\rho(\hat{x}^k) \leq O\left(\frac{L_d R_d^2}{k}\right) + O(\epsilon_{in})$
- Primal suboptimality: $|f(\hat{z}^k) - f^*| \leq O\left(\frac{L_d R_d}{k}\right) + O(\epsilon_{in})$
- Primal feasibility violation: $\|A\hat{z}^k - b\| \leq O\left(\frac{L_d R_d}{k}\right) + O\left(\sqrt{\frac{\epsilon_{in}}{k}}\right)$

Here $R_d = \min_{x^* \in X^*} \|x^0 - x^*\|$. For a desired outer accuracy $\epsilon_{out}$ we can choose:

$$k_{out} = \left\lceil \frac{L_d R_d^2}{\epsilon_{out}} \right\rceil \quad \& \quad \epsilon_{in} = O(\epsilon_{out})$$

\[\downarrow\]

$|f(\hat{z}^{k_{out}}) - f^*| \leq O(\epsilon_{out}) \quad \& \quad \|A\hat{z}^{k_{out}} - b\| \leq O(\epsilon_{out})$
Inexact dual fast gradient augm. Lagrangian alg.

IDFGAL

\[
\begin{align*}
    x^k &= \left[ y^k + \frac{1}{L_d} \nabla \tilde{d}_\rho(y^k) \right] + \\
    y^{k+1} &= x^k + \frac{\theta_k - 1}{\theta_{k+1}} (x^k - x^{k-1})
\end{align*}
\]

- \( \nabla \tilde{d}_\rho(x^k) = Az^k - b \) inexact dual gradient
- recall \( z^k = \bar{z}(x^k) \) approximate solution of the inner problem for given \( x^k \)
- \( \theta_{k+1} = \frac{1+\sqrt{4\theta_k^2+1}}{2} \) with \( \theta^0 = 1 \)

Define the average dual and primal sequence \((\hat{x}^k, \hat{z}^k)\):

\[
\begin{align*}
    \hat{x}^k &= \frac{\sum_{j=0}^{k} \theta_j x^j}{S^k_\theta} \\
    \hat{z}^k &= \frac{\sum_{j=0}^{k} \theta_j \bar{z}^j}{S^k_\theta}, \quad \text{with} \quad S^k_\theta = \sum_{j=0}^{k} \theta_j
\end{align*}
\]
Convergence rate - estimates for suboptimality/infeas.

**Main result:** estimates for suboptimality and feasibility violation for the approximate primal and dual solutions $\hat{z}^k$ and $\hat{x}^k$ in average generated by (IDFGAL)

**Theorem 8:**

- **Dual suboptimality:** $f^* - d_\rho(\hat{x}^k) \leq O\left(\frac{L_d R^2}{k^2}\right) + O\left(k\right) \epsilon_{in}$
- **Primal suboptimality:** $|f(\hat{z}^k) - f^*| \leq O\left(\frac{L_d R^2}{k^2}\right) + O\left(k\right) \epsilon_{in}$
- **Primal feasibility violation:** $\|A\hat{z}^k - b\| \leq O\left(\frac{L_d R^2}{k^2}\right) + O\left(\sqrt{\frac{\epsilon_{in}}{k}}\right)$

For a desired outer accuracy $\epsilon_{out}$ we can choose:

$$k_{out} = \left\lfloor 2R_d \sqrt{\frac{L_d}{\epsilon_{out}}} \right\rfloor \quad \& \quad \epsilon_{in} = O(\epsilon_{out} \sqrt{\epsilon_{out}})$$

↓

$$|f(\hat{z}^{k_{out}}) - f^*| \leq O(\epsilon_{out}) \quad \& \quad \|A\hat{z}^{k_{out}} - b\| \leq O(\epsilon_{out})$$
Numerical results VIII

→ our estimates are dependent on $x^*$ via $R_d = \|x^0 - x^*\|$...BUT there are no good upper bounds on $x^*$ (e.g. [Nesterov’12] and [Richter’13])!

→ average number of outer iterations (theoretical and real) for $\epsilon_{out} = 10^{-3}$, $\rho = 1$ and variable horizon length $N$ on MPC problems with 10 random initial state

![Graphs showing outer iterations vs prediction horizon N for Algorithm IDGM and Algorithm IDFGM](image_url)
Numerical results IX

→ number of outer iterations for $\epsilon_{\text{out}} = 10^{-3}$, $\rho = 1$, $N = 20$ and variable inner accuracy $\epsilon_{\text{in}}$ for one QP problem
Numerical results X

→ random QP problems:

\[
\min_{lb \leq z \leq ub} \left\{ 0.5z^T Q z + q^T z : \text{s.t. } Az = b \right\}
\]

→ \( Q \in \mathbb{R}^{r \times n} \) si \( A \in \mathbb{R}^{\left\lceil \frac{n}{2} \right\rceil \times n} \) generated randomly from an uniform distribution, with zero mean and unit covariance

→ \( Q \leftarrow Q^T Q \), with \( \text{rang}(Q) \) varying between 0.5\( n \) and 0.9\( n \)

→ \( ub = -lb = 1 \) and \( b \) random

→ for each \( n \) ⇒ 10 QP problems

→ comparison with other QP solvers: quadprog (Matlab R2008b), Sedumi 1.3 (C++), Cplex 12.4 (IBM ILOG)(C++) and Gurobi 5.0.1(C++).
Numerical results X

→ average time for $\epsilon_{\text{out}} = 10^{-3}$ and $\rho$ chosen variable according to some rule [Boyd’11]
Inexact dual - dual aug.: common & different features

\[ \min_{z \in Z : Az = b} f(z) \]

- \( f \) (nonsmooth) convex, \( Z \) simple
- dual augmented Lagrangian formulation
- dual aug. function - Lipschitz gradient
- inexact solution of inner
- \( \mathcal{L}_\rho(\bar{z}(x), x) - \mathcal{L}_\rho(z(x), x) \leq \epsilon_{in}^2 \)
- inexact dual (fast) gradient \( \Rightarrow \)
  \[ \epsilon_{in} \approx \epsilon_{out} (\epsilon_{in} \approx \epsilon_{out} \sqrt{\epsilon_{out}}) \]
- complete estimates on suboptimality/infeasibility
- drawback: no tight upper bounds on \( R_d = \|x^*\| \) [Nesterov’12]

\[ \min_{u \in U : g(u) \leq 0} f(u) \]

- \( f \) strongly convex, \( U \) simple
- dual Lagrangian formulation
- dual function - Lipschitz gradient
- inexact solution of inner
- \( \mathcal{L}(\bar{u}(x), x) - \mathcal{L}(u(x), x) \leq \epsilon_{in} \)
- inexact dual (fast) gradient \( \Rightarrow \)
  \[ \epsilon_{in} \approx \epsilon_{out} (\epsilon_{in} \approx \epsilon_{out} \sqrt{\epsilon_{out}}) \]
- complete estimates on suboptimality/infeasibility
- drawback: no tight upper bounds on \( R_d = \|x^*\| \) [Nedich’09], [Patrinos’14]
Conclusions

- motivation: embedded/distributed MPC & many other engineering applications
- all these problems recast as convex optimization problems
- solve using (augmented) dual formulation & notion of inexact dual gradient
- our analysis is based on Lipschitz/error bound property of the dual
- our analysis uses primal last iterate/average of iterates
- analyze (inexact) dual first order (augmented Lagrangian) algorithms
- derive complete rate analysis for the proposed algorithms
- tight estimates on primal suboptimality/feasibility violation

Future work:

- faster methods & improve the existing results on convergence rate (e.g. (DFG) converges linearly under error bound?)
- existing upper bounds on $\|x^*\|$ are NOT tight & effects of inexact arithmetics!
- dual methods do NOT guarantee feasibility/stability! $\Rightarrow$ constraint tightening?

...Optimizers are not (yet...) out of job...!
Talk based on papers


http://acse.pub.ro/person/ion-necoara/