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

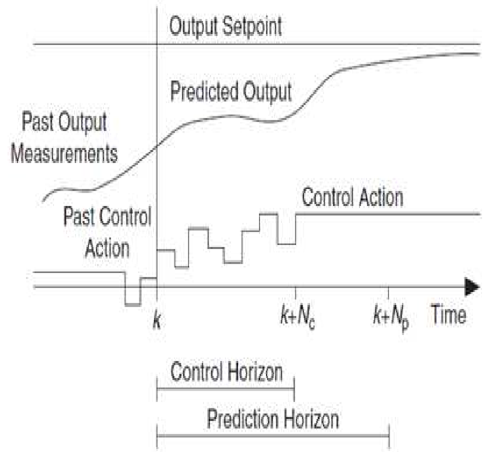
$$\Downarrow$$

Embedded Model Predictive Control (MPC)

- Linear systems: $x_{t+1} = A_x x_t + B_u u_t$
- State/input constraints: $x_t \in X \;\&\; u_t \in U$ (X, U simple sets, e.g. box)
- Stage/final costs: $\ell(x,u) \;\&\; \ell_{\rm f}(x)$ (e.g. quadratic)
- Finite horizon optimal control of length N:

$$\min_{x_t \in X, u_t \in U} \sum_{t=0}^{N-1} \ell(x_t, u_t) + \ell_{\rm f}(x_N)$$

$$\text{s.t. :} \quad x_{t+1} = A_x x_t + B_u u_t, \; x_0 = x$$



Optimization problem formulation

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

$$z = \left[x_1^T \cdots x_N^T \; u_0^T \cdots u_{N-1}^T\right]^T \in \mathbb{R}^n \quad \& \quad Z = \prod_{t=1}^{N-1} X \times X_\mathrm{f} \times \prod_{t=1}^{N} U$$

$$f(z) = \sum_{t=0}^{N-1} \ell(x_t, u_t) + \ell_\mathrm{f}(x_N)$$

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

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

$$\text{s.t.: } Az = b, \quad z \in Z,$$

✓ 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

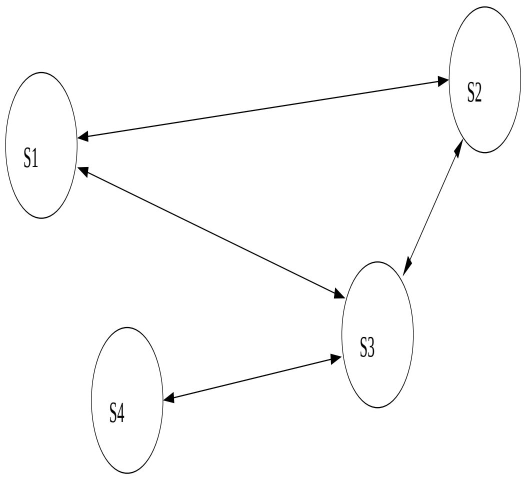
Distributed control requires:

- 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: $\ell^i(x^i, u^i)$ & $\ell^i_{\mathrm{f}}(x^i)$
- Finite horizon optimal control of length N:



$$\min_{x^i_t \in X^i, u^i_t \in U^i} \quad \sum_i \sum_t \ell^i(x^i_t, u^i_t) + \ell^i_{\mathrm{f}}(x^i_N)$$
$$\text{s.t. :} \quad x^i_{t+1} = \sum_{j \in \mathcal{N}^i} A^{ij}_x x^j_t + B^{ij}_u u^j_t, \; x_0 = x$$

Centralized optimization problem formulation

✓ Dense formulation of centralized MPC (i.e. elimination of states via dynamics):

✓ Define input trajectories for each subsystem:

$$\mathbf{u}_i = [u_0^i \cdots u_{N-1}^i] \quad \& \quad \mathbf{u} = [\mathbf{u}_1 \cdots \mathbf{u}_M]$$

$$f(\mathbf{u}) = \sum_i \sum_t \ell^i(x_t^i, u_t^i) + \ell_f^i(x_N^i)$$

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

$$f^* = \min_{\mathbf{u}_i \in \mathbf{U}^i} f(\mathbf{u}_1, \cdots, \mathbf{u}_M) \qquad \Longleftrightarrow \qquad \min_{\mathbf{u} \in \mathbf{U}} f(\mathbf{u})$$

$$\text{s.t.}: \; g(\mathbf{u}_1, \cdots, \mathbf{u}_M) \leq 0 \qquad \qquad \text{s.t.}: \; g(\mathbf{u}) \leq 0$$

✓ 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 $\mathcal{S}$
- 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: $\mathcal{U}_r(u_r)$

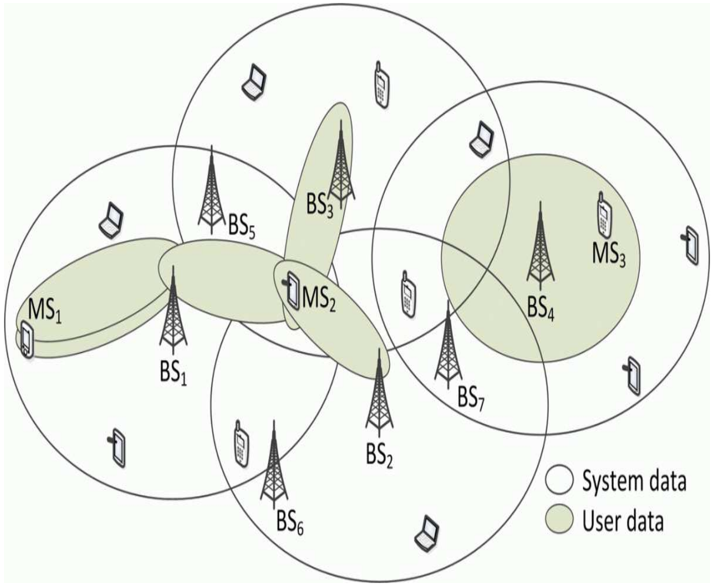
$$\max_{u_r \geq 0} \sum_{r \in \mathcal{S}} \mathcal{U}_r(u_r)$$

$$\text{s.t.}: \sum_{r: l \in r} u_r \leq c_l \quad \forall l \in \mathcal{L}$$

$$\Downarrow$$

$$\min_{\mathbf{u} \in \mathbf{U}} f(\mathbf{u})$$

$$\text{s.t.}: \underbrace{G\mathbf{u} + g}_{g(\mathbf{u})} \leq 0$$



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

$$\min_{\mathbf{u}\in\mathbf{U}} f(\mathbf{u}) \qquad \& \qquad \min_{\mathbf{u}\in\mathbf{U}} f(\mathbf{u})$$

$$\text{s.t. : } A\mathbf{u} = b \qquad\qquad \text{s.t. : } g(\mathbf{u}) \leq 0$$

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

✓ "Simplest" first order method: Gradient Method

solutie ec. $F(u) = 0 \Longleftarrow$ 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

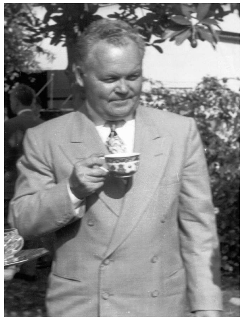
  - First derived by Cauchy (1847)
  - Cauchy solved a nonlinear system of 6 equations with 6 unknowns

A. Cauchy. *Methode generale pour la resolution des systemes d'equations simultanees*, C. R. Acad. Sci. Paris, 25, 1847

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. rezults (≈ 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 α_k can be chosen as: constant, Wolfe conditions, backtracking, ideal,...
- Advantages of GM:
    - reduced complexity per iteration - $\mathcal{O}(n) +$ 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)

[Dev:14] Devolder, Glineur, Nesterov, *First-order methods of smooth convex optimization with inexact oracle*, Math. Prog., 2014

[Nec:14] Necoara, Nedelcu, *Rate analysis of inexact dual first order methods: application to dual decomposition*, IEEE T. Automatic Control, 2014

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 σ, 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 $\Longrightarrow$ develope new methods with better performance:

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

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

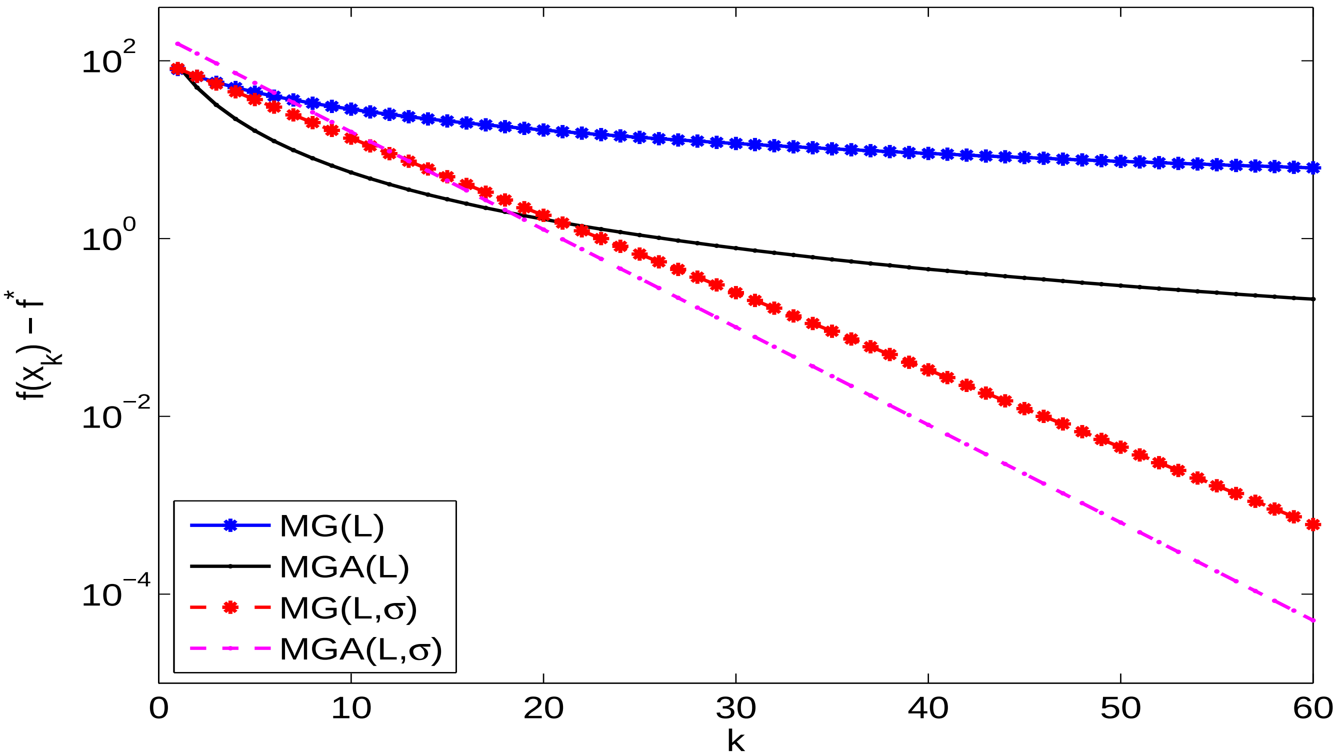
Linear combination: $v_{k+1} = u_{k+1} + \theta_k(u_{k+1} - u_k)$

- Initial points $u_0 = v_0$
- θ_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!).

[Nes:83] Nesterov, *A method for unconstrained convex minimization problem with the rate of convergence* $\mathcal{O}(1/k^2)$, Soviet. Math. Docl., 269, 1983.

Convergence GM versus FGM

Conditions f	**GM**	**FGM**
$f(u)$ convex & $\nabla f(u)$ Lipschitz	$\mathcal{O}\left(\frac{LR^2}{k}\right)$	$\mathcal{O}\left(\frac{LR^2}{k^2}\right)$
$\nabla f(u)$ Lipschitz & $f(u)$ strong convex	$\mathcal{O}\left(\left(\frac{L-\sigma}{L+\sigma}\right)^k\right)$	$\mathcal{O}\left(\left(1-\sqrt{\frac{\sigma}{L}}\right)^k\right)$



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

$$\boxed{u_{k+1} = [u_k - \alpha_k \nabla f(u_k)]_U}$$

- 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 $Gu + g \leq 0 \quad \& \quad Au = b$, with G and A general matrices), then projection is difficult $\rightarrow$ dual approach

Gradient type methods: primal vrs. dual

- Dual approach: solve dual problem

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

- 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!

- [Nec:13] Necoara, Nedelcu, *Rate analysis of inexact dual first order methods: application to dual decomposition*, IEEE T. Automatic Control, 2014
- [Nec:14] Nedelcu, Necoara, Dinh, *Computational complexity of inexact gradient augmented Lagrangian methods: application to constrained MPC*, SIAM J. Control & Optimization
- [LuoTse:93] Luo, Tseng,, *On the convergence rate of dual ascent methods for strictly convex minimization*, Math. Oper. Res., 18, 1993

PART I

Inequality constrained problems - dual formulation

$$f^* = \min_{\mathbf{u} \in \mathbf{U}} \{ f(\mathbf{u}) : \; g(\mathbf{u}) \leq 0 \}$$

Primal convex problem, where:

- f strong convex (e.g. $\nabla^2 f \succeq \sigma_{\text{f}} I_n$)
- g convex & $\|\nabla g\|_F \leq c_g$ (e.g. $g(u) = G\mathbf{u} + g$)
- **U** simple set (e.g. box, ball,...) & strong duality holds

GOAL - for desired accuracy ϵ_{out} compute nr. iterations k and generate $\hat{\mathbf{u}}^k \in \mathbf{U}$:

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

Approach - use the dual formulation

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

• Define $\mathbf{u}(x)$ solution of **inner problem**: $\mathbf{u}(x) \in \arg\min_{\mathbf{u} \in \mathbf{U}} \mathcal{L}(\mathbf{u}, x)$, 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

$$f^* = \min_{\mathbf{u}\in\mathbf{U}}\{f(\mathbf{u}):\ g(\mathbf{u})\leq 0\} \quad \Leftrightarrow \quad \max_{x\in\mathbb{R}^m_+} d(x)$$

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

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

$$\Downarrow$$

Lemma 1 Dual function $d(x)$ satisfies:

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

Lemma 2 Dual function $d(x)$ and inner solution $\mathbf{u}(x)$ satisfy:

- primal suboptimality 1: $\frac{\sigma_{\mathrm{f}}}{2}\|\mathbf{u}(x) - \mathbf{u}^*\|^2 \leq f^* - d(x) \qquad \forall x \geq 0$
- primal suboptimality 2: $|f(\mathbf{u}(x)) - f^*| \leq c_g\, (\|x - x^*\| + \|x^*\|)\, \|\mathbf{u}(x) - \mathbf{u}^*\|$
- primal feasibility: $\|[g(\mathbf{u}(x))]_+\| \leq c_g \|\mathbf{u}(x) - \mathbf{u}^*\|$

Dual first order methods

$$f^* = \min_{\mathbf{u}\in\mathbf{U}} \{f(\mathbf{u}) : \ g(\mathbf{u}) \leq 0\} \quad \Leftrightarrow \quad \max_{x\in\mathbb{R}^m_+} d(x)$$

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

$$d(x) + \langle \nabla d(x), y - x\rangle - \frac{L_{\rm d}}{2}\|y-x\|^2 \leq d(y) \qquad \forall x, y \in \mathbb{R}^m_+$$

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. $x^k = \left[y^k + \alpha_k \nabla d(y^k)\right]_+$
3. $y^{k+1} = x^k + \frac{\theta_k - 1}{\theta_{k+1}}(x^k - x^{k-1})$

- **Dual Gradient (DG)**: parameters (step size) $\frac{1}{L_{\rm d}} \leq \alpha_k \leq \frac{1}{L_{\rm D}}$ and $\theta_k = 1$
- **Dual Fast Gradient (DFG)**: parameters $\alpha_k = \frac{1}{L_{\rm d}}$ and $\theta_{k+1} = \frac{1+\sqrt{1+4\theta_k^2}}{2}$

Dual gradient alg. (DG)

Dual Gradient (DG): chooses parameters $\frac{1}{L_{\mathrm{d}}} \leq \alpha_k \leq \frac{1}{L_{\mathrm{D}}}$ and $\theta_k = 1$ in Alg. (DFO)

(DG) : $$x^{k+1} = \left[x^k + \alpha_k \nabla d(x^k)\right]_+$$

- dual gradient: $\nabla d(x^k) = g(\mathbf{u}^k)$
- $\mathbf{u}^k = \mathbf{u}(x^k)$ exact solution of the inner problem for given x^k, i.e.

$$\mathbf{u}^k = \arg\min_{\mathbf{u} \in \mathbf{U}} \mathcal{L}(\mathbf{u}, x^k)$$

✓ Define dual and primal last iterate sequences:

$$(x^k \quad \& \quad \mathbf{u}^k)$$

✓ Define dual and primal average sequences:

$$\left(x^k \qquad \& \qquad \hat{\mathbf{u}}^k = \frac{\sum_{j=0}^k \alpha_j \mathbf{u}^j}{S^k} \quad \text{with} \quad S^k = \sum_{j=0}^k \alpha_j\right)$$

Convergence rate - estimates for suboptimality/infeas.

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

- Dual suboptimality: $f^* - d(x^k) \leq \mathcal{O}\left(\frac{L_\mathrm{d} R_\mathrm{d}^2}{k}\right)$

Theorem 1: for approximate primal solution in average $\hat{\mathbf{u}}^k$

- Primal suboptimality: $|f(\hat{\mathbf{u}}^k) - f^*| \leq \mathcal{O}\left(\frac{L_\mathrm{d} R_\mathrm{d}^2}{k}\right)$
- Primal feasibility violation: $\|[h(\hat{\mathbf{u}}^k)]_+\| \leq \mathcal{O}\left(\frac{L_\mathrm{d} R_\mathrm{d}}{k}\right)$

Theorem 2: for approximate primal solution in last iterate $\mathbf{u}^k$

- Primal suboptimality: $|f(\mathbf{u}^k) - f^*| \leq \mathcal{O}\left(\frac{L_\mathrm{d} R_\mathrm{d}^2}{\sqrt{k}}\right)$
- Primal feasibility violation: $\|[h(\mathbf{u}^k)]_+\| \leq \mathcal{O}\left(\frac{L_\mathrm{d} R_\mathrm{d}}{\sqrt{k}}\right)$

Here $R_\mathrm{d} = \min_{x^* \in X^*} \|x^0 - x^*\|$.

Dual fast gradient alg. (DFG)

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

DFG

$$\begin{cases} x^k & = \left[y^k + \frac{1}{L_\mathrm{d}} \nabla d(y^k) \right]_+ \\ y^{k+1} & = x^k + \frac{\theta_k - 1}{\theta_{k+1}} (x^k - x^{k-1}) \end{cases}$$

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

✓ Define dual and primal last iterate sequences:

$$\left(x^k \quad \& \quad \mathbf{v}^k = \mathbf{u}(x^k) = \arg\min_{\mathbf{u}\in\mathbf{U}} \mathcal{L}(\mathbf{u}, x^k) \right)$$

✓ Define dual and primal average sequences:

$$\left(x^k \qquad \& \qquad \hat{\mathbf{u}}^k = \frac{\sum_{j=0}^k \theta_j \mathbf{u}^j}{S_\theta^k} \quad \text{with} \quad S_\theta^k = \sum_{j=0}^k \theta_j \right)$$

Convergence rate - estimates for suboptimality/infeas.

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

- Dual suboptimality: $f^* - d(x^k) \leq \mathcal{O}\left(\frac{L_{\mathrm{d}} R_{\mathrm{d}}^2}{k^2}\right)$

Theorem 3: for approximate primal solution in average $\hat{\mathbf{u}}^k$

- Primal suboptimality: $|f(\hat{\mathbf{u}}^k) - f^*| \leq \mathcal{O}\left(\frac{L_{\mathrm{d}} R_{\mathrm{d}}^2}{k^2}\right)$
- Primal feasibility violation: $\|[g(\hat{\mathbf{u}}^k)]_+\| \leq \mathcal{O}\left(\frac{L_{\mathrm{d}} R_{\mathrm{d}}}{k^2}\right)$

Theorem 4: for approximate primal solution in last iterate $\mathbf{v}^k$

- Primal suboptimality: $|f(\mathbf{v}^k) - f^*| \leq \mathcal{O}\left(\frac{L_{\mathrm{d}} R_{\mathrm{d}}^2}{k}\right)$
- Primal feasibility violation: $\|[g(\mathbf{v}^k)]_+\| \leq \mathcal{O}\left(\frac{L_{\mathrm{d}} R_{\mathrm{d}}}{k}\right)$

Similarly, $R_{\mathrm{d}} = \min_{x^* \in X^*} \|x^0 - x^*\|$.

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: $L_\text{d} = \frac{c_g^2}{\sigma_\text{f}}$

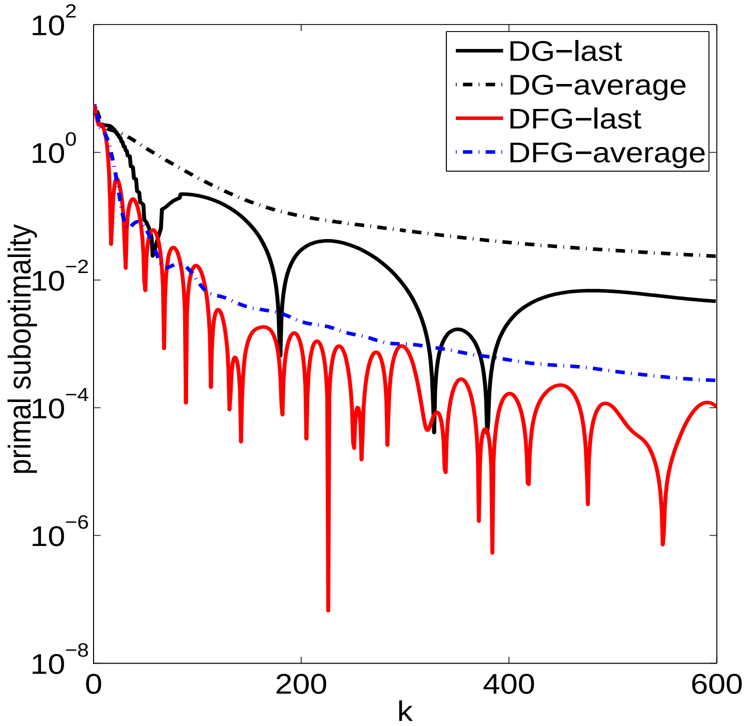
✓ Summary of estimates of primal suboptimality & infeasibility for (DFO):

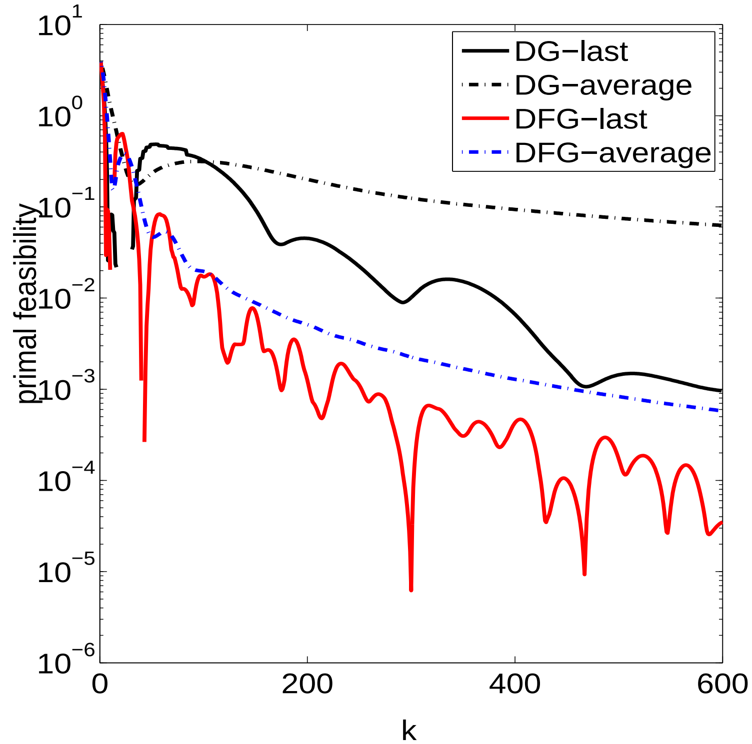
Alg.	**DG**	**DFG**
last	$\mathcal{O}\left(\frac{1}{\sqrt{k}}\right)$	$\mathcal{O}\left(\frac{1}{k}\right)$
average	$\mathcal{O}\left(\frac{1}{k}\right)$	$\mathcal{O}\left(\frac{1}{k^2}\right)$

✓ 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

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

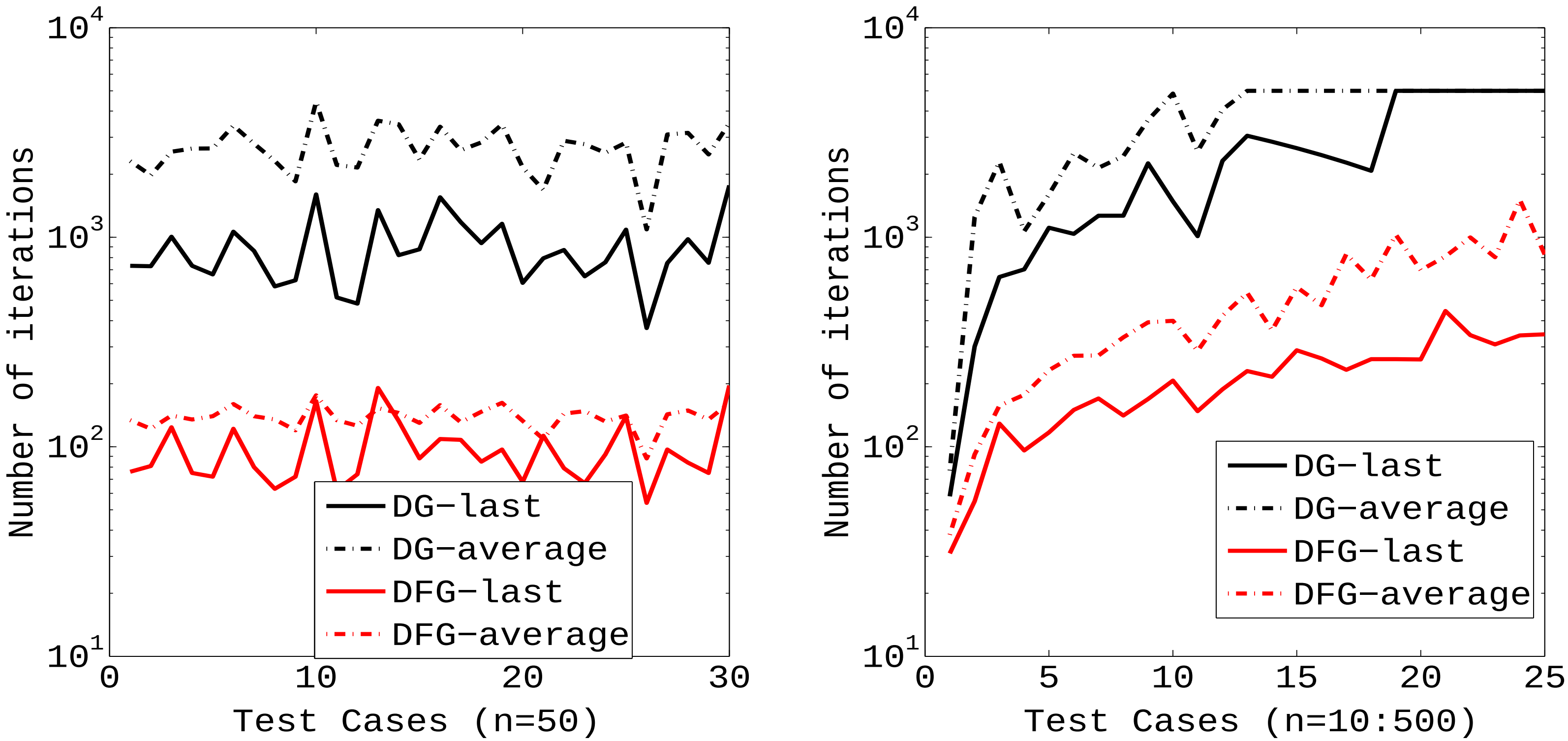




- typical behavior of (DFO) methods: oscillating in primal feasibility & suboptimality!
- dual first order methods performs better in last iterates than in average sequences (worst case analysis says differently)!

Numerical results II

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



- 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

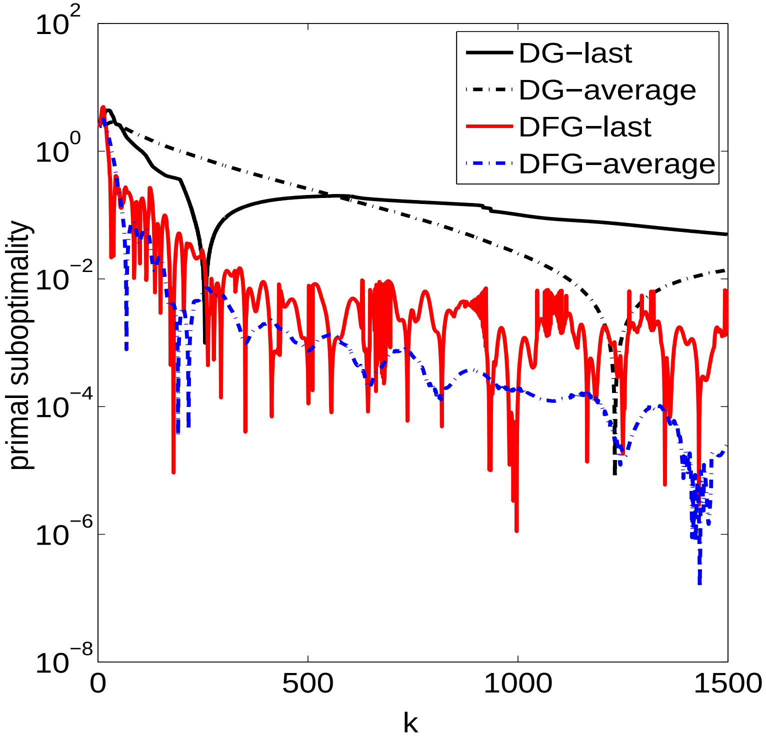
$$\min_{\text{lb} \leq \mathbf{u} \leq \text{ub},\ G\mathbf{u}+g \leq 0} \mathbf{u}^T Q \mathbf{u} + q^T \mathbf{u} + \gamma \log(1 + \exp^{a^T \mathbf{u}}), \qquad G \in \mathbb{R}^{1.5n \times n},\ \sigma = 1,\ \gamma = 0.1$$

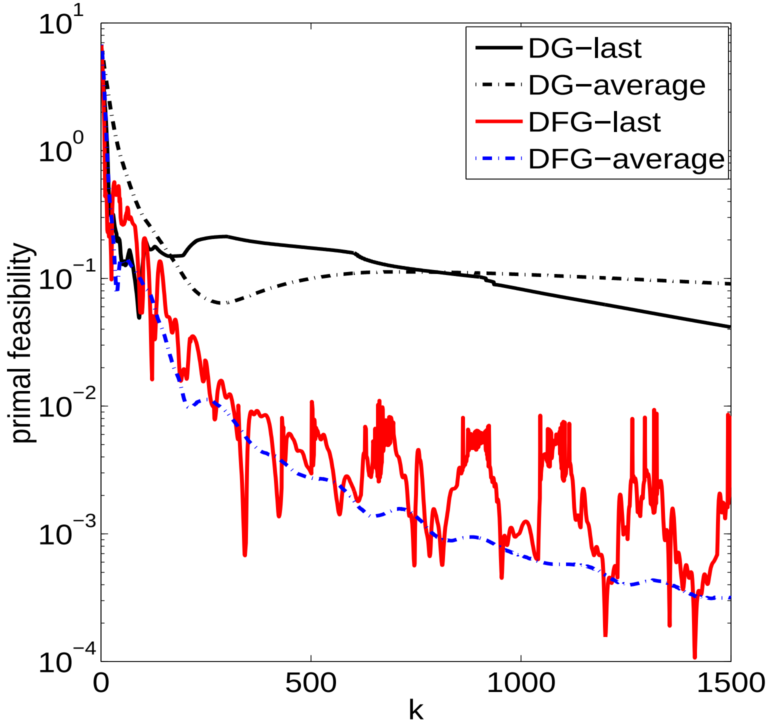
Alg./n	10	50	10^2	10^3	$5 * 10^3$	10^4
$k_{\text{last}}^{\text{DG}}$	35	195	463	782	1.147	2.155
$k_{\text{avg.}}^{\text{DG}}$	527	3.423	12.697	–	–	–
k_{last}^{DFG}	19	61	97	198	276	292
$k_{\text{avg.}}^{DFG}$	41	108	186	381	563	582

- average results for 10 random problems ($\epsilon = 10^{-2}$)
- Q and G sparse for large problems (few tens ≈ 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

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





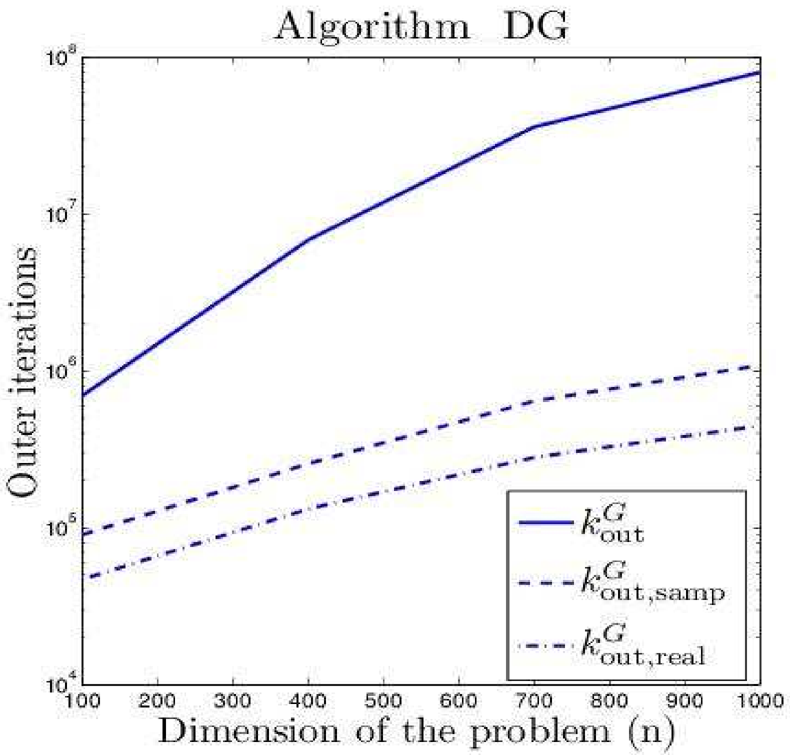
- practical performance usually better in the last iterate
- however, there are also cases where the practical performance is comparable with the theoretical estimates

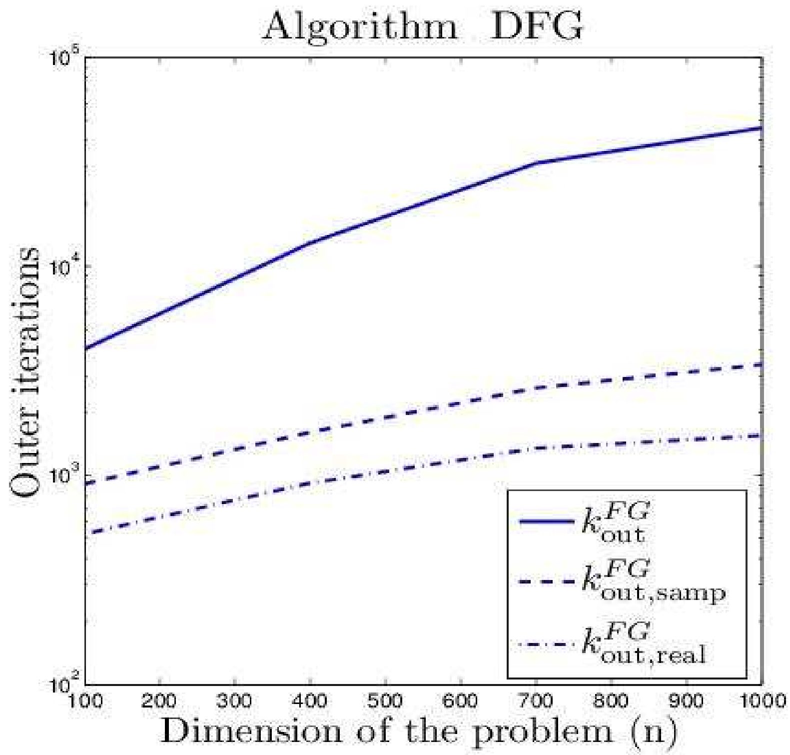
Numerical results V

→ our estimates are dependent on x^* via $R_\text{d} = \|x^0 - x^*\|$...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

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

- 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

$$\bar{\mathbf{u}}(x) \approx \arg\min_{\mathbf{u} \in \mathbf{U}} f(\mathbf{u}) + \langle x, g(\mathbf{u}) \rangle$$

such that the following stopping criterion holds

$$\bar{\mathbf{u}}(x) \in \mathbf{U}, \;\; \mathcal{L}\left(\bar{\mathbf{u}}(x), x\right) - \mathcal{L}\left(\mathbf{u}(x), x\right) \leq \epsilon_{\text{in}}/3$$

- Introduce two notions: inexact dual function and gradient
  $\bar{d}(x) = \mathcal{L}\left(\bar{\mathbf{u}}(x), x\right)$ & $\nabla \bar{d}(x) = g(\bar{\mathbf{u}}(x))$

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 $x, y \in \mathbb{R}^m_+$:

$$\bar{d}(x) + \left\langle \nabla \bar{d}(x), y - x \right\rangle - L_{\text{d}} \|y - x\|^2 - \epsilon_{\text{in}}$$

$$\leq d(y) \leq$$

$$\leq \bar{d}(x) + \left\langle \nabla \bar{d}(x), y - x \right\rangle$$

Convergence rate of (DFO) under inexact dual gradients

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

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

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

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

- Dual suboptimality: $f^* - d(x^k) \leq \mathcal{O}\left(\frac{L_{\text{d}} R_{\text{d}}^2}{k^2}\right) + k\epsilon_{\text{in}}$
- Primal suboptimality: $|f(\hat{z}^k) - f^*| \leq \mathcal{O}\left(\frac{L_{\text{d}} R_{\text{d}}^2}{k^2}\right) + k\epsilon_{\text{in}}$
- Primal feasibility violation: $\|[h(\hat{\mathbf{u}}^k)]_+\| \leq \mathcal{O}\left(\frac{L_{\text{d}} R_{\text{d}}}{k^2}\right) + \mathcal{O}\left(\sqrt{\frac{\epsilon_{\text{in}}}{k}}\right)$

Convergence rate - estimates for suboptimality/infeas.

For a desired outer accuracy ϵ_{out} we can choose in *inexact* (DFO) alg.:

Alg.	inexact (DG)	inexact (DFG)
average	$k_{\text{out}} = \mathcal{O}\left(\left\lceil \frac{L_{\text{d}} R_{\text{d}}^2}{\epsilon_{\text{out}}} \right\rceil\right)$ & $\epsilon_{\text{in}} = \epsilon_{\text{out}}$	$k_{\text{out}} = \mathcal{O}\left(\left\lceil R_{\text{d}} \sqrt{\frac{L_{\text{d}}}{\epsilon_{\text{out}}}} \right\rceil\right)$ & $\epsilon_{\text{in}} = \epsilon_{\text{out}} \sqrt{\epsilon_{\text{out}}}$

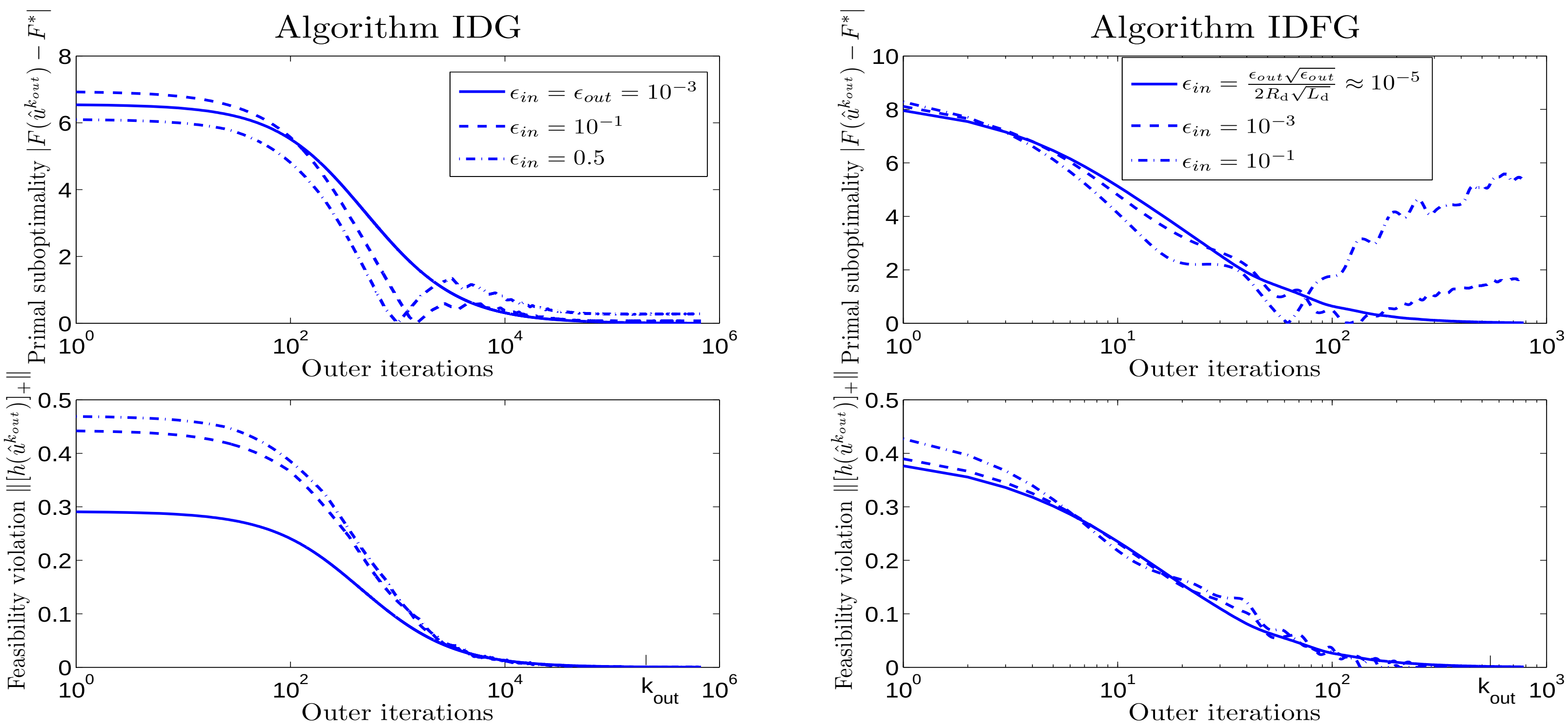
$$\Downarrow$$

$$|f(\hat{\mathbf{u}}^{k_{\text{out}}}) - f^*| \leq \epsilon_{\text{out}} \qquad \& \qquad \|[g(\hat{\mathbf{u}}^{k_{\text{out}}})]_+\| \leq \epsilon_{\text{out}}$$

$$\Downarrow$$

(DFG) more sensitive to errors (inexact gradients) than (DG)

Numerical results VI



- nr. outer iterations on random QP for $\epsilon_{\text{out}} = 10^{-3}$ & inner accuracy ϵ_{in} variable
- (DFG) better than (DG) in average primal sequence ($\mathcal{O}(\frac{1}{k^2})$ instead of $\mathcal{O}(\frac{1}{k})$)
- 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:

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

- When we have linear convergence of first order methods for above primal problem? $\Rightarrow$ 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)? $\Rightarrow$ Answer: error bound type property
- In practice: strong convexity for f is not found often in applications

Example 1:

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

Example 2:

$$f(x) = \log\left(1 + e^{a^T x}\right) \Longrightarrow L = \|\nabla^2 f(x)\| = \|\frac{e^{a^T x}}{(1+e^{a^T x})^2} a a^T\| \leq \frac{\|a\|^2}{4} \quad \& \quad \sigma = 0$$

Issues on strong convexity for dual function d

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

$$\min_{\mathbf{u}:\ G\mathbf{u}+g\leq 0} f(\mathbf{u}) \qquad \& \qquad \mathbf{U} = \mathbb{R}^n$$

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. $d(x) = \min_{\mathbf{u}\in\mathbb{R}^n} f(\mathbf{u}) + \langle x, G\mathbf{u} + g\rangle$ 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**: $d(x) = \min_{\mathbf{u}\in\mathbb{R}^n} f(\mathbf{u}) + \langle x, G\mathbf{u}\rangle = -\tilde{f}(-G^T x)$. Here $\tilde{f}$ is conjugate function of f and is strong convex provided that f has Lipschitz gradient!

Linear convergence of dual gradient method

$$(P): \quad \min_{\mathbf{u}\in\mathbb{R}^n:\ G\mathbf{u}+g\leq 0} f(\mathbf{u})$$

Assume f is strong convex & Lipschitz gradient

- Primal first order methods converge linearly!
- f strong convex, then dual fct. $d(x) = \min_{\mathbf{u}\in\mathbb{R}^n} f(\mathbf{u}) + \langle x, G\mathbf{u}+g\rangle$ 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: $X = \{x:\ Cx \leq c\}$
- Note: corresponding dual problem of (P) has these assumptions!

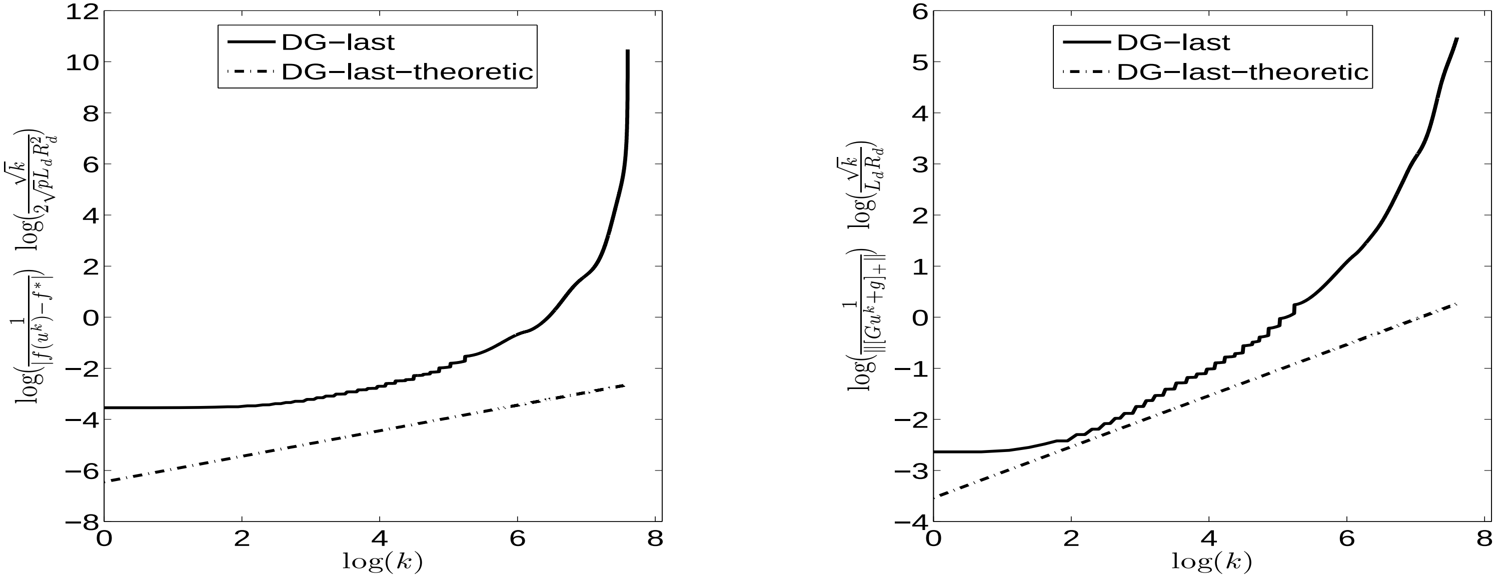
SATISFY AN ERROR BOUND PROPERTY

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

- [NecNed:14] Necoara, Nedelcu, *On linear convergence of a distributed dual gradient algorithm for linearly constrained separable convex problems*, Automatica, partially accepted, 2014

Numerical results VII

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



- 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 $\mathcal{O}(\frac{1}{\sqrt{k}})$
- 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

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

GOAL - for desired accuracy ϵ_{out} compute $k_{\text{out}}, \epsilon_{\text{in}}$ and ϵ_c and generate $\hat{\mathbf{u}}^{k_{\text{out}}} \in \mathbf{U}$:

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

$$\Downarrow$$

Slater vector approach

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

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

Stability: follows from stability of suboptimal MPC with quadratic cost by choosing ϵ_{out} adequately:

$$\Downarrow$$

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

where x^+ is the next state in the MPC scheme and $\tilde{\mathbf{u}}^+$ 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 ϵ_{out} compute nr. iterations k and generate $\hat{z}^k \in Z$:

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

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

- 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**: $z(x) \in \arg\min_{z \in Z} \mathcal{L}(z, x)$, 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:

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

✓ Function d_ρ 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 ∇d_ρ Lipschitz with $L_{\mathrm{d}} = \frac{1}{\rho}$

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

- BUT gradient $\nabla d_\rho(x)$ requires EXACT solution of inner problem ⇒ hard to compute
- Possible remedy: approximate solution of inner problem ⇒ 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

$$\bar{z}(x) \in Z, \;\; \mathcal{L}_\rho\left(\bar{z}(x), x\right) - \mathcal{L}_\rho\left(z(x), x\right) \le \mathcal{O}(\epsilon_{\text{in}}^2)$$

or

$$\bar{z}(x) \in Z, \; \left\langle \nabla \mathcal{L}_\rho\left(\bar{z}(x), x\right), z - \bar{z}(x)\right\rangle \ge -\epsilon_{\text{in}} \; \forall z \in Z$$

Define two notions:

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

Lemma: $\|\nabla \bar{d}_\rho(x) - \nabla d_\rho(x)\| \le \mathcal{O}(\epsilon_{\text{in}})$

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

$$\bar{d}_\rho(x) + \left\langle \nabla \bar{d}_\rho(x), y - x\right\rangle - \frac{L_{\text{d}}}{2}\|y - x\|^2 - \epsilon_{\text{in}}$$

$$\le d_\rho(y) \le$$

$$\le \bar{d}_\rho(x) + \left\langle \nabla \bar{d}_\rho(x), y - x\right\rangle$$

Inexact dual gradient augm. Lagrangian alg.

From previous descent type lemma $\Rightarrow$ smooth dual $\Rightarrow$ dual first order methods

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

where

- $\nabla \bar{d}_\rho(x^k) = A\bar{z}^k - b$ inexact dual gradient
- $\bar{z}^k = \bar{z}(x^k)$ approximate solution of the inner problem for given x^k
$$\bar{z}^k \approx \arg\min_{z \in Z} f(z) + \langle x^k, Az - b \rangle + \frac{\rho}{2}\|Az - b\|^2$$
- REMARK: the theory works also for $\epsilon_{\text{in}} = 0$ (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 $(\hat{x}^k, \hat{z}^k)$:

$$\left(\hat{x}^k = \frac{\sum_{j=0}^k \alpha^j x^{j+1}}{S^k} \quad \& \quad \hat{z}^k = \frac{\sum_{j=0}^k \alpha^j \bar{z}^j}{S^k}, \quad \text{with } S^k = \sum_{j=0}^k \alpha^j \right)$$

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

Here $R_\mathrm{d} = \min_{x^* \in X^*} \|x^0 - x^*\|$. For a desired outer accuracy ϵ_out we can choose:

$$k_\mathrm{out} = \left\lfloor \frac{L_\mathrm{d} R_\mathrm{d}^2}{\epsilon_\mathrm{out}} \right\rfloor \quad \& \quad \epsilon_\mathrm{in} = \mathcal{O}(\epsilon_\mathrm{out})$$

$$\Downarrow$$

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

Inexact dual fast gradient augm. Lagrangian alg.

IDFGAL

$$\begin{cases} x^k & = \left[y^k + \frac{1}{L_{\rm d}} \nabla \bar{d}_\rho(y^k) \right]_+ \\ y^{k+1} & = x^k + \frac{\theta_k - 1}{\theta_{k+1}} (x^k - x^{k-1}) \end{cases}$$

- $\nabla \bar{d}_\rho(x^k) = A\bar{z}^k - b$ inexact dual gradient
- recall $\bar{z}^k = \bar{z}(x^k)$ 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 $(\hat{x}^k, \hat{z}^k)$:

$$\left(\hat{x}^k = \frac{\sum_{j=0}^k \theta_j x^j}{S_\theta^k} \quad \& \quad \hat{z}^k = \frac{\sum_{j=0}^k \theta_j \bar{z}^j}{S_\theta^k}, \quad \text{with } S_\theta^k = \sum_{j=0}^k \theta_j \right)$$

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: $f^* - d_\rho(\hat{x}^k) \leq \mathcal{O}\left(\frac{L_\text{d} R_\text{d}^2}{k^2}\right) + \mathcal{O}(k)\,\epsilon_\text{in}$
- Primal suboptimality: $|f(\hat{z}^k) - f^*| \leq \mathcal{O}\left(\frac{L_\text{d} R_\text{d}}{k^2}\right) + \mathcal{O}(k)\,\epsilon_\text{in}$
- Primal feasibility violation: $\|A\hat{z}^k - b\| \leq \mathcal{O}\left(\frac{L_\text{d} R_\text{d}}{k^2}\right) + \mathcal{O}\left(\sqrt{\frac{\epsilon_\text{in}}{k}}\right)$

For a desired outer accuracy ϵ_out we can choose:

$$k_\text{out} = \left\lfloor 2R_\text{d}\sqrt{\frac{L_\text{d}}{\epsilon_\text{out}}} \right\rfloor \quad \& \quad \epsilon_\text{in} = \mathcal{O}(\epsilon_\text{out}\sqrt{\epsilon_\text{out}})$$

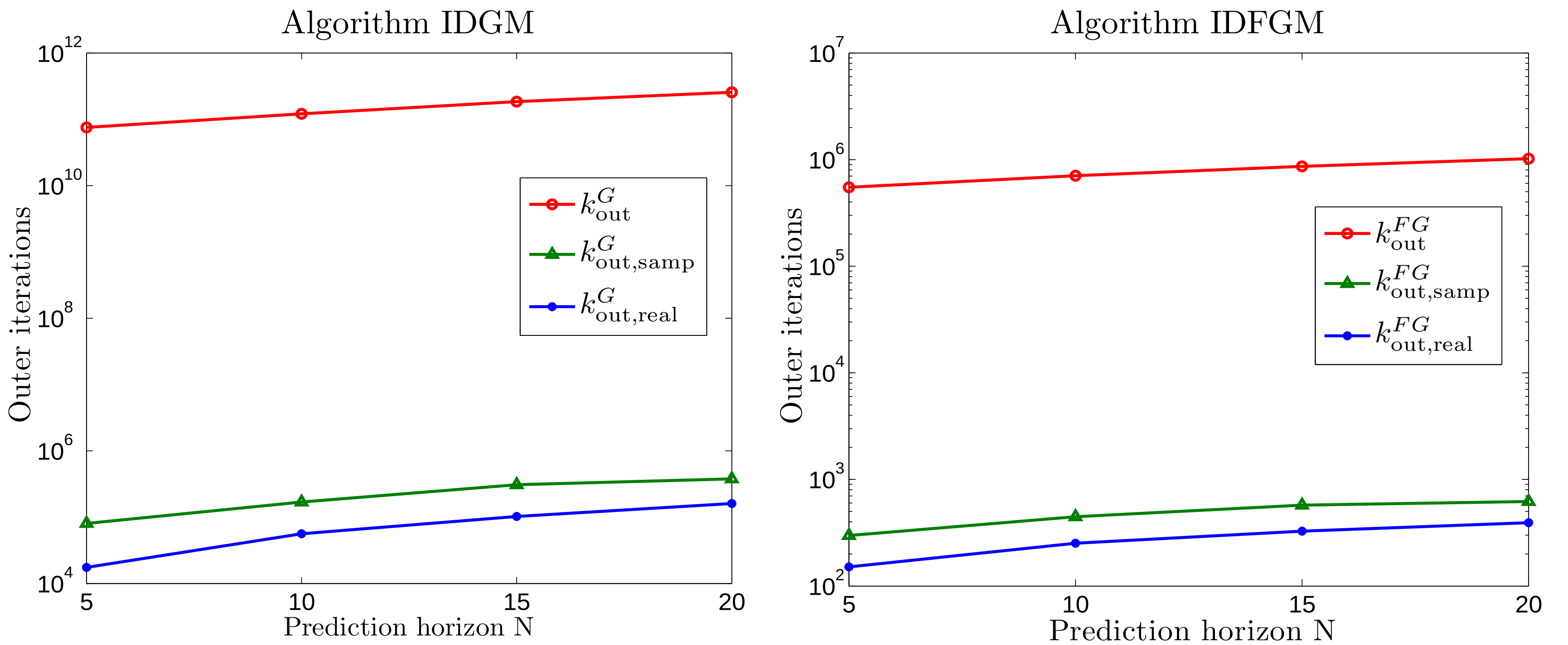
$$\Downarrow$$

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

Numerical results VIII

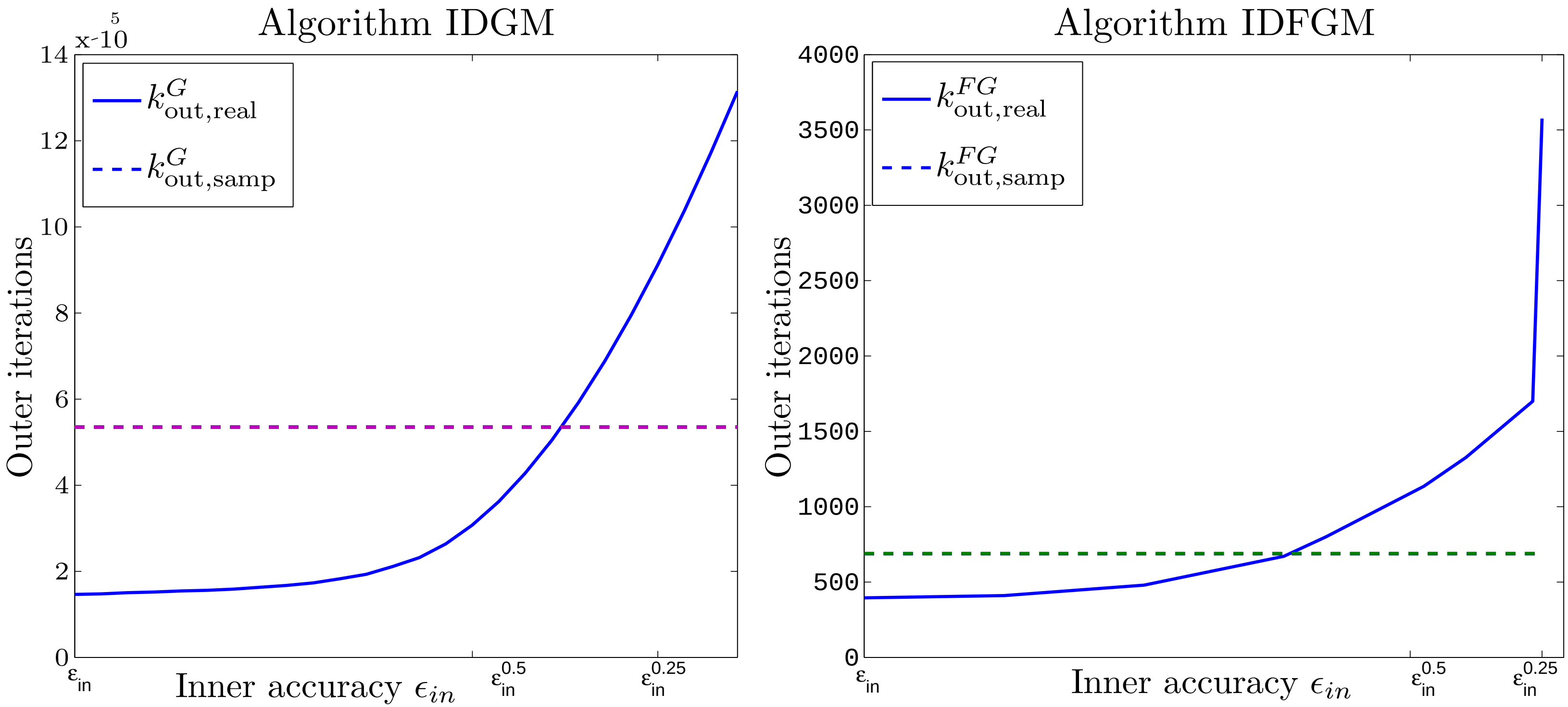
$\rightarrow$ our estimates are dependent on x^* via $R_{\text{d}} = \|x^0 - x^*\|$...BUT there are no good upper bounds on x^* (e.g. [Nesterov'12] and [Richter'13])!

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



Numerical results IX

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



Numerical results X

$\rightarrow$ random QP problems:

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

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

$\rightarrow Q \leftarrow Q^T Q$, with rang$(Q)$ varying between $0.5n$ and $0.9n$

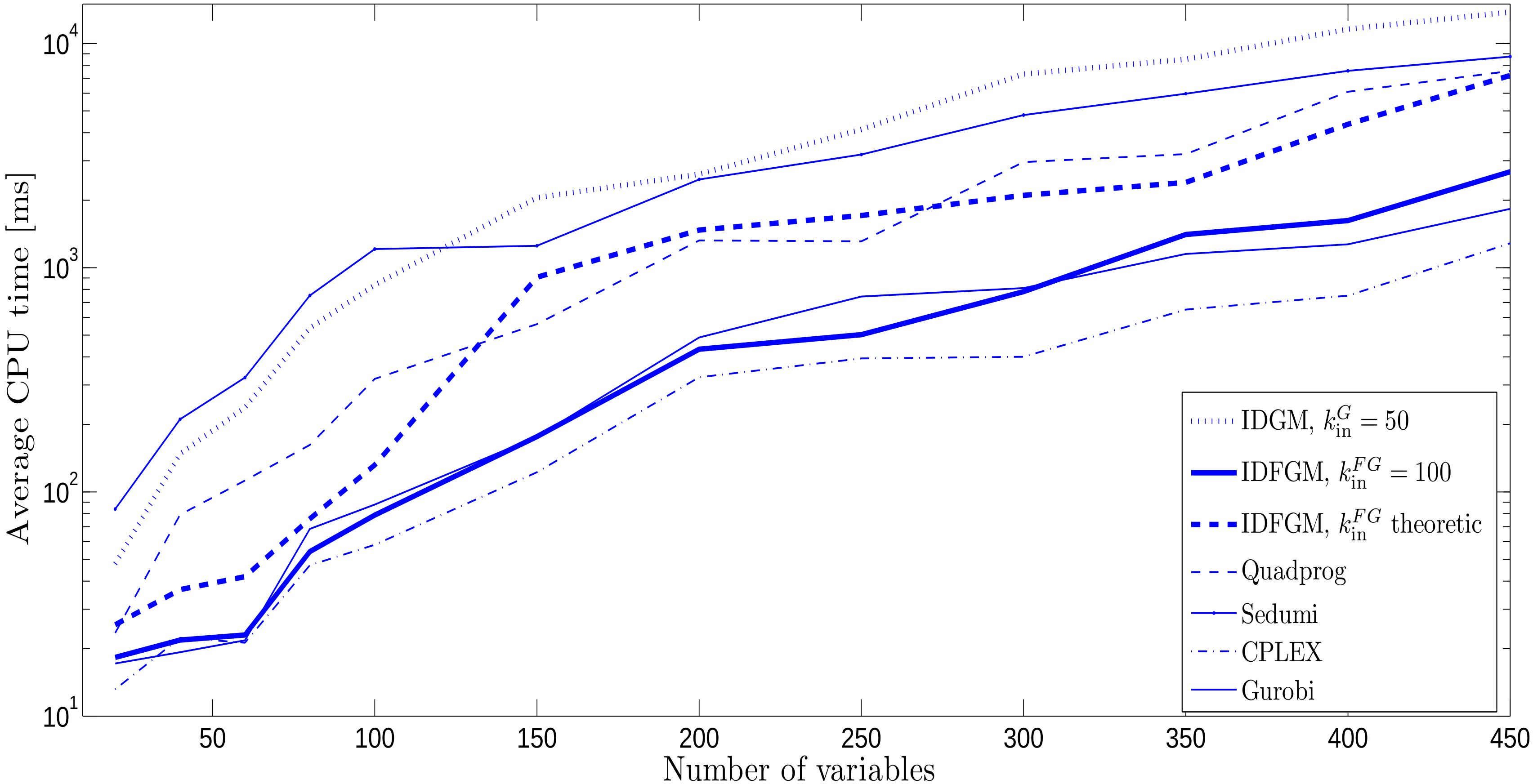
$\rightarrow \mathrm{ub} = -\mathrm{lb} = 1$ and b random

$\rightarrow$ for each $n \Rightarrow 10$ QP problems

$\rightarrow$ 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

$\rightarrow$ average time for $\epsilon_{\text{out}} = 10^{-3}$ and ρ chosen variable according to some rule [Boyd'11]



Inexact dual - dual aug.: common $\&$ different features

Lets go together

Lets go to get her

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

- $\min_{\mathbf{u} \in \mathbf{U} : g(\mathbf{u}) \leq 0} f(\mathbf{u})$
- f strongly convex, **U** simple
- dual Lagrangian formulation
- dual function - Lipschitz gradient
- inexact solution of inner
- $\mathcal{L}(\bar{\mathbf{u}}(x), x) - \mathcal{L}(\mathbf{u}(x), x) \leq \epsilon_{\text{in}}$
- inexact dual (fast) gradient $\Rightarrow$ $\epsilon_{\text{in}} \approx \epsilon_{\text{out}} (\epsilon_{\text{in}} \approx \epsilon_{\text{out}} \sqrt{\epsilon_{\text{out}}})$
- complete estimates on suboptimality/infeasibility
- drawback: no tight upper bounds on $R_{\text{d}} = \|x^*\|$ [Nedich'09],[Patrinos'14]

Conclusions

- motivation: embedded/distributed MPC & many other engineering applications
- all these problem 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

- Necoara, Nedelcu, *Rate analysis of inexact dual first order methods: Application to distributed MPC for network systems*, IEEE T. Automatic Control, 2014.
- Nedelcu, Necoara, Tran-Dinh, *Computational Complexity of Inexact Gradient Augmented Lagrangian Methods: Application to Constrained MPC*, SIAM J. Control & Optimization, 2014.
- Necoara, Suykens, *Application of a smoothing technique to decomposition in convex optimization*, IEEE Trans. Automatic Control, 2008
- Necoara, Nedelcu, *On linear convergence of a distributed dual gradient alg. for linearly constrained separable convex problems*, partially acc. Automatica, 2014
- Necoara, *Computational complexity certification for embedded MPC based on dual gradient method*, submitted to Systems & Control Lett., 2014
- Necoara, Keviczky, *An adaptive constraint tightening approach to linear MPC based on approximation alg. for optimization*, J. Opt. Control Appl. & Met., 2014.

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