A Coordination-Decomposition Algorithm for Solving Distributed Non-convex Programs

Jean-Hubert Hours

Colin N. Jones

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Motivation & Challenges

Algorithm Description

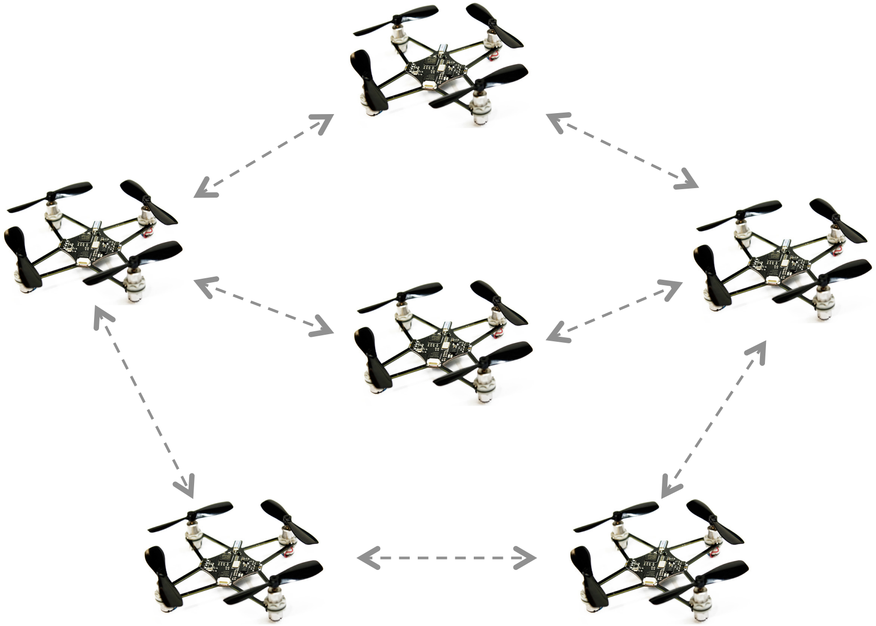
Convergence Analysis

Numerical Examples

Motivation

- Agents collaborate to achieve a common control objective

Power grids



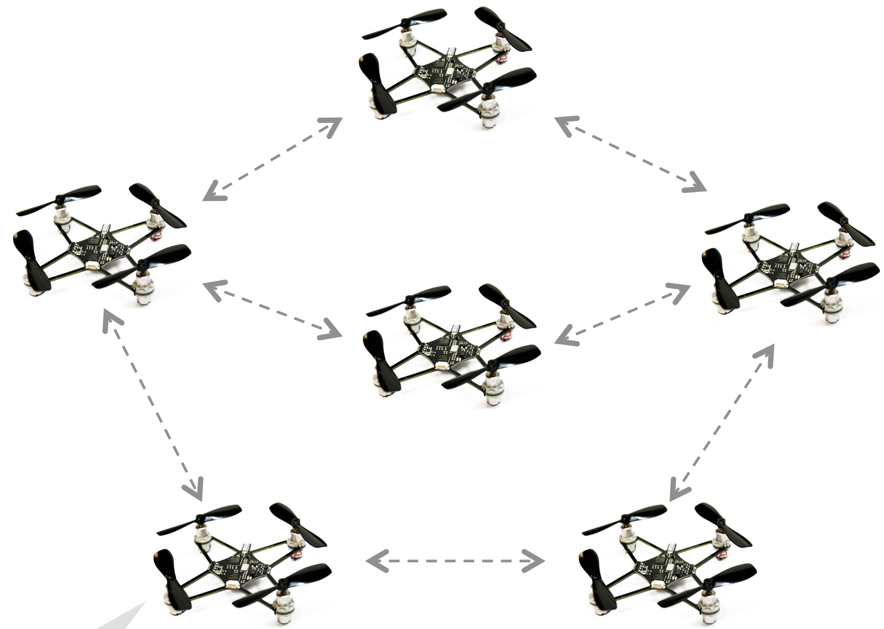
Formation Stabilisation

- **Goal:** Decentralise / Parallelise computations for solving nonlinear optimal control / NMPC problems

Motivation

- Agents collaborate to achieve a common control objective

Power grids



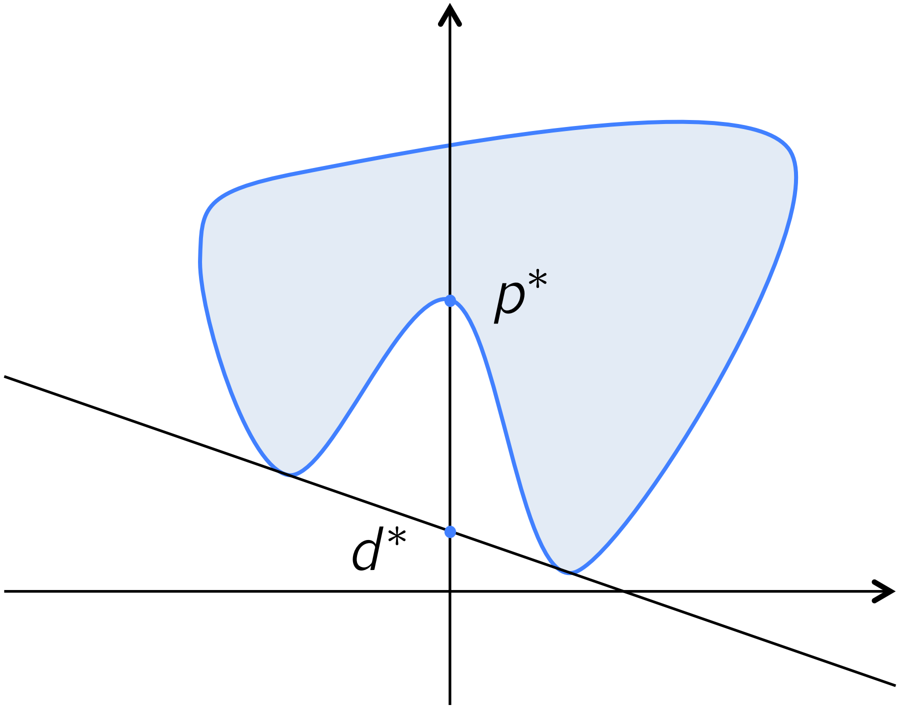
Formation Stabilisation

Highly nonlinear dynamics

- **Goal:** Decentralise / Parallelise computations for solving nonlinear optimal control / NMPC problems

Challenges

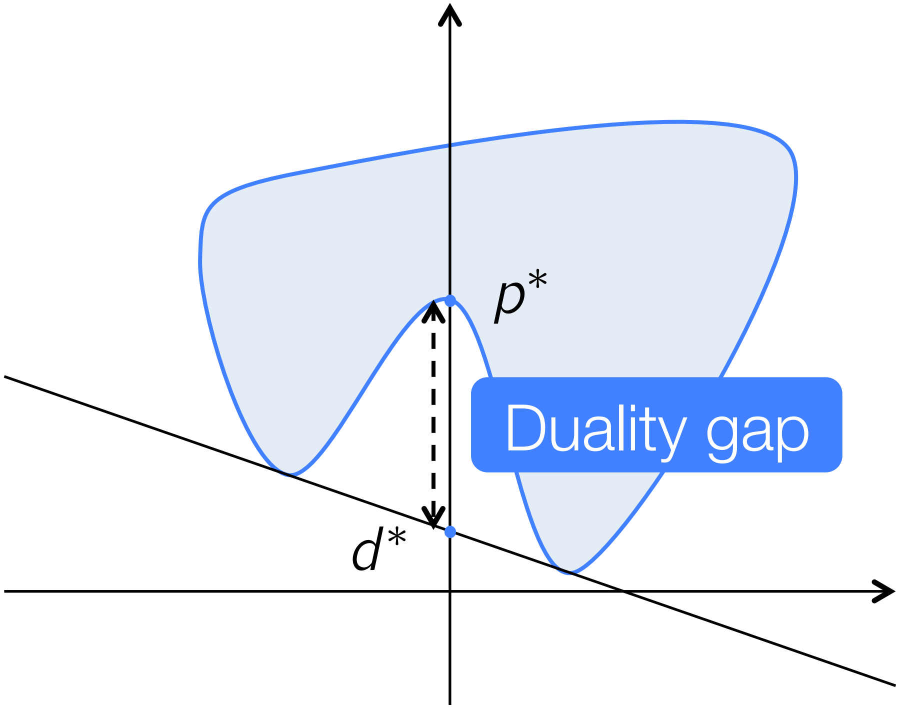
- Distributed optimisation generally addressed via Lagrangian decomposition…but in a non-convex setting



- Convergence of block-coordinate descent (BCD) not clear in non-convex cases

Challenges

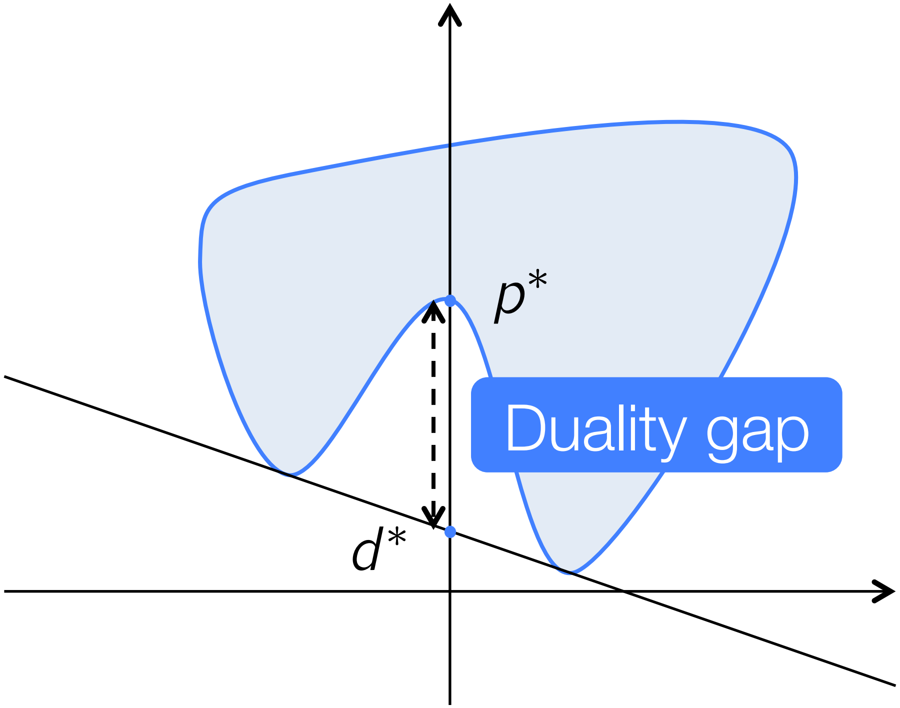
- Distributed optimisation generally addressed via Lagrangian decomposition…but in a non-convex setting



- Convergence of block-coordinate descent (BCD) not clear in non-convex cases

Challenges

- Distributed optimisation generally addressed via Lagrangian decomposition…but in a non-convex setting



- Convergence of block-coordinate descent (BCD) not clear in non-convex cases

$\Longrightarrow$ Sequential Convex Programming ?

Motivation & Challenges

Algorithm Description

Convergence Analysis

Numerical Examples

Scope: Distributed Semi-algebraic Programs

$$\underset{z_1,\ldots,z_N}{\text{minimise}} \sum_{i=1}^{N} f_i(z_i) + G(z_1,\ldots,z_N)$$

s.t.

$$H(z_1,\ldots,z_N) = 0$$

$$g_i(z_i) = 0$$

$$z_i \in \mathcal{Z}_i,\ i \in \{1,\ldots,N\}$$

- Twice continuously differentiable semi-algebraic functions
- Constraint sets $\mathcal{Z}_i$ closed semi-algebraic convex
- Second order optimality
- Linear independence constraint qualification

Scope: Distributed Semi-algebraic Programs

$$\underset{z_1,\dots,z_N}{\text{minimise}} \sum_{i=1}^{N} f_i(z_i) + \boxed{G(z_1,\dots,z_N)}$$

s.t.

$$\boxed{H(z_1,\dots,z_N) = 0}$$

$$g_i(z_i) = 0$$

$$z_i \in \mathcal{Z}_i,\ i \in \{1,\dots,N\}$$

Cost coupling

Constraint coupling

- Twice continuously differentiable **semi-algebraic** functions
- Constraint sets $\mathcal{Z}_i$ closed **semi-algebraic** convex
- Second order optimality
- Linear independence constraint qualification

The Augmented Lagrangian Framework

Inner primal loop :

$$z^k \approx \operatorname{argmin} \; G(z) + \left(\nu^k + \frac{\rho^k}{2} H(z)\right)^\top H(z)$$
$$+ \sum_{i=1}^{N} f_i(z_i) + \left(\mu_i^k + \frac{\rho^k}{2} g_i(z_i)\right)^\top g_i(z_i) + \delta_{\mathcal{Z}_i}(z_i)$$

Outer dual / penalty loop :

$$\mu_i^{k+1} = \mu_i^k + \rho^k g(z_i^k), \; i \in \{1, \ldots, N\}$$
$$\nu^{k+1} = \nu^k + \rho^k H(z^k)$$
$$\rho^{k+1} \leftarrow \alpha \rho^k, \; \alpha > 1$$

The Augmented Lagrangian Framework

<u>Inner primal loop</u> :

$$z^k \approx \operatorname{argmin}\ G(z) + \left(\boldsymbol{\nu}^k + \frac{\boldsymbol{\rho}^k}{2}\boldsymbol{H}(z)\right)^\top \boldsymbol{H}(z)$$

$$+ \sum_{i=1}^{N} f_i(z_i) + \left(\boldsymbol{\mu}_i^k + \frac{\boldsymbol{\rho}^k}{2}\boldsymbol{g}_i(z_i)\right)^\top \boldsymbol{g}_i(z_i) + \delta_{\mathcal{Z}_i}(z_i)$$

<u>Outer dual / penalty loop</u> :

$$\mu_i^{k+1} = \mu_i^k + \rho^k g(z_i^k),\ i \in \{1, \dots, N\}$$

$$\nu^{k+1} = \nu^k + \rho^k H(z^k)$$

$$\rho^{k+1} \leftarrow \alpha\rho^k,\ \alpha > 1$$

The Augmented Lagrangian Framework

Inner primal loop :

$$z^k \approx \operatorname{argmin} \; G(z) + \left(\nu^k + \frac{\rho^k}{2} H(z)\right)^\top H(z)$$
$$+ \sum_{i=1}^{N} f_i(z_i) + \left(\mu_i^k + \frac{\rho^k}{2} g_i(z_i)\right)^\top g_i(z_i) + \delta_{\mathcal{Z}_i}(z_i)$$

Outer dual / penalty loop :

$$\mu_i^{k+1} = \mu_i^k + \rho^k g(z_i^k), \; i \in \{1, \dots, N\}$$
$$\nu^{k+1} = \nu^k + \rho^k H(z^k)$$
$$\rho^{k+1} \leftarrow \alpha \rho^k, \; \alpha > 1$$

The Augmented Lagrangian Framework

<u>Inner primal loop</u> : Smooth non-separable

$$z^k \approx \operatorname{argmin} \; \overbrace{G(z) + \left(\nu^k + \frac{\rho^k}{2} H(z)\right)^\top H(z)}^{\text{Smooth non-separable}}$$

$$+ \underbrace{\sum_{i=1}^{N} f_i(z_i) + \left(\mu_i^k + \frac{\rho^k}{2} g_i(z_i)\right)^\top g_i(z_i) + \delta_{\mathcal{Z}_i}(z_i)}_{\text{Non-smooth separable}}$$

Non-smooth separable

<u>Outer dual / penalty loop</u> :

$$\mu_i^{k+1} = \mu_i^k + \rho^k g(z_i^k), \; i \in \{1, \ldots, N\}$$

$$\nu^{k+1} = \nu^k + \rho^k H(z^k)$$

$$\rho^{k+1} \leftarrow \alpha \rho^k, \; \alpha > 1$$

The Augmented Lagrangian Framework

In theory:

- Local convergence to KKT if "sufficient" criticality in the primal
- Globalisation by updating dual when "sufficient" feasibility

In practice:

- Decomposition among agents in the primal program
- Coordination via dual updates

Issue: Quadratic penalty term $\frac{\rho^k}{2} \|H(z)\|_2^2$ non-separable, even though $H(z)$ separable

What if Convex ?

$$\underset{x,y}{\text{minimise}} \; f(x) + g(y)$$
$$\text{s.t.} \; Ax - y = 0$$

ADMM:

$$x^{k+1} = \underset{x}{\operatorname{argmin}} \, f(x) + \left(\mu^k\right)^\top (Ax - y^k) + \frac{\rho}{2} \left\|Ax - y^k\right\|_2^2$$

$$y^{k+1} = \operatorname{argmin} \, g(y) + \left(\mu^k\right)^\top (Ax^{k+1} - y) + \frac{\rho}{2} \left\|Ax^{k+1} - y\right\|_2^2$$

$$\mu^{k+1} = \mu^k + \rho\left(Ax^{k+1} - y^{k+1}\right)$$

What if Convex ?

$$\underset{x,y}{\text{minimise}} \;\; f(x) + g(y)$$
$$\text{s.t.} \;\; Ax - y = 0$$

ADMM:

$$x^{k+1} = \underset{x}{\operatorname{argmin}} \, f(x) + \left(\mu^k\right)^\top (Ax - y^k) + \frac{\rho}{2} \|Ax - y^k\|_2^2$$
$$y^{k+1} = \operatorname{argmin} g(y) + \left(\mu^k\right)^\top (Ax^{k+1} - y) + \frac{\rho}{2} \left\|Ax^{k+1} - y\right\|_2^2$$
$$\mu^{k+1} = \mu^k + \rho\left(Ax^{k+1} - y^{k+1}\right)$$

Main idea: Some form of Block Coordinate Descent in the primal

Proximal Alternating Linearised Minimisations

- Decomposition of primal functional $L_{\rho^k}\left(z_1, \ldots, z_N, \mu^k\right)$

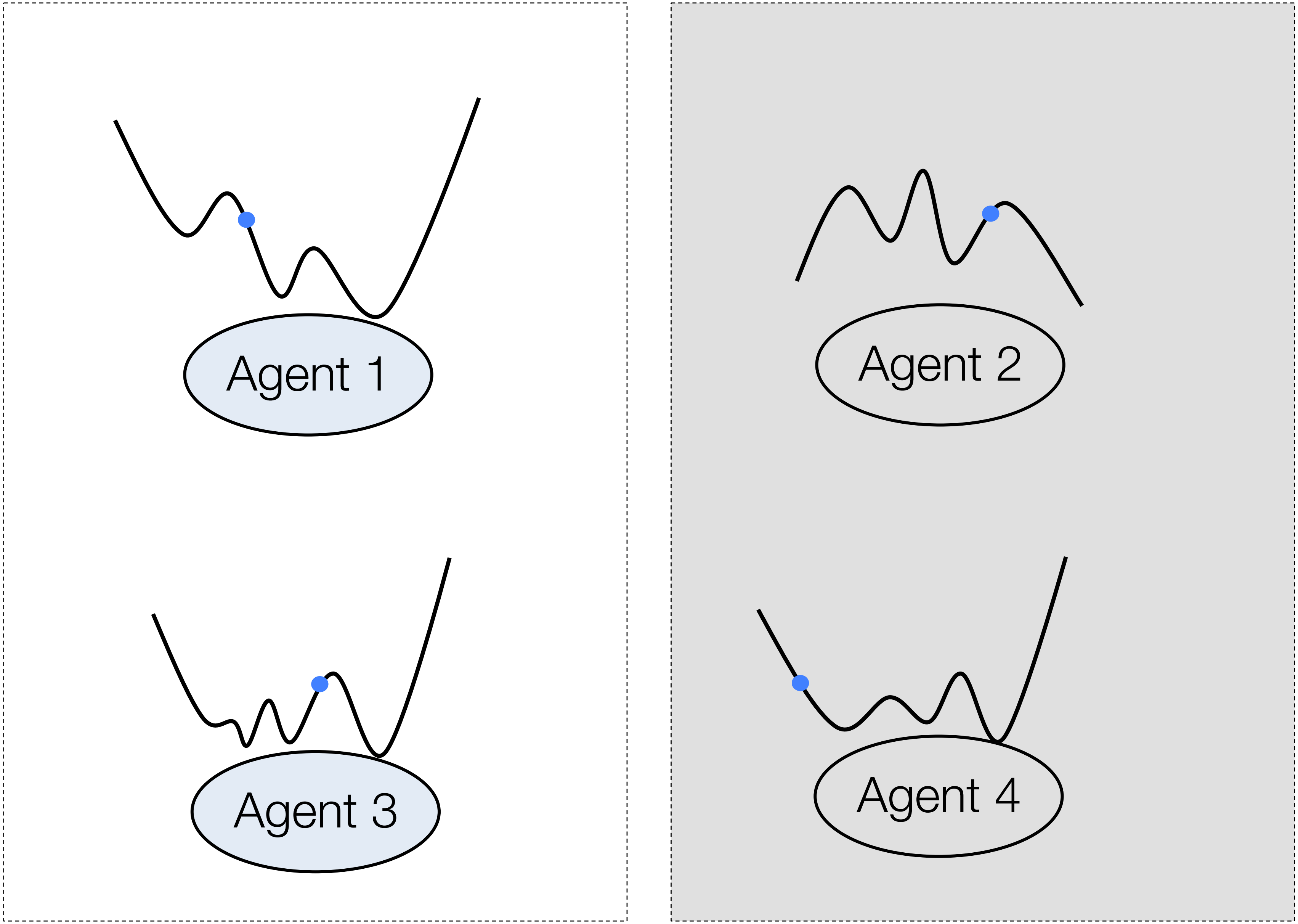
$$z_i^{l+1} = \underset{z_i \in \mathcal{Z}_i}{\operatorname{argmin}} \; \nabla_i L_{\rho^k}\left(z_1^{l+1}, \ldots, z_N^l, \mu^k\right)^\top \left(z_i - z_i^l\right) + \frac{c_i^l}{2} \left\| z_i - z_i^l \right\|_2^2$$

$$\Updownarrow$$

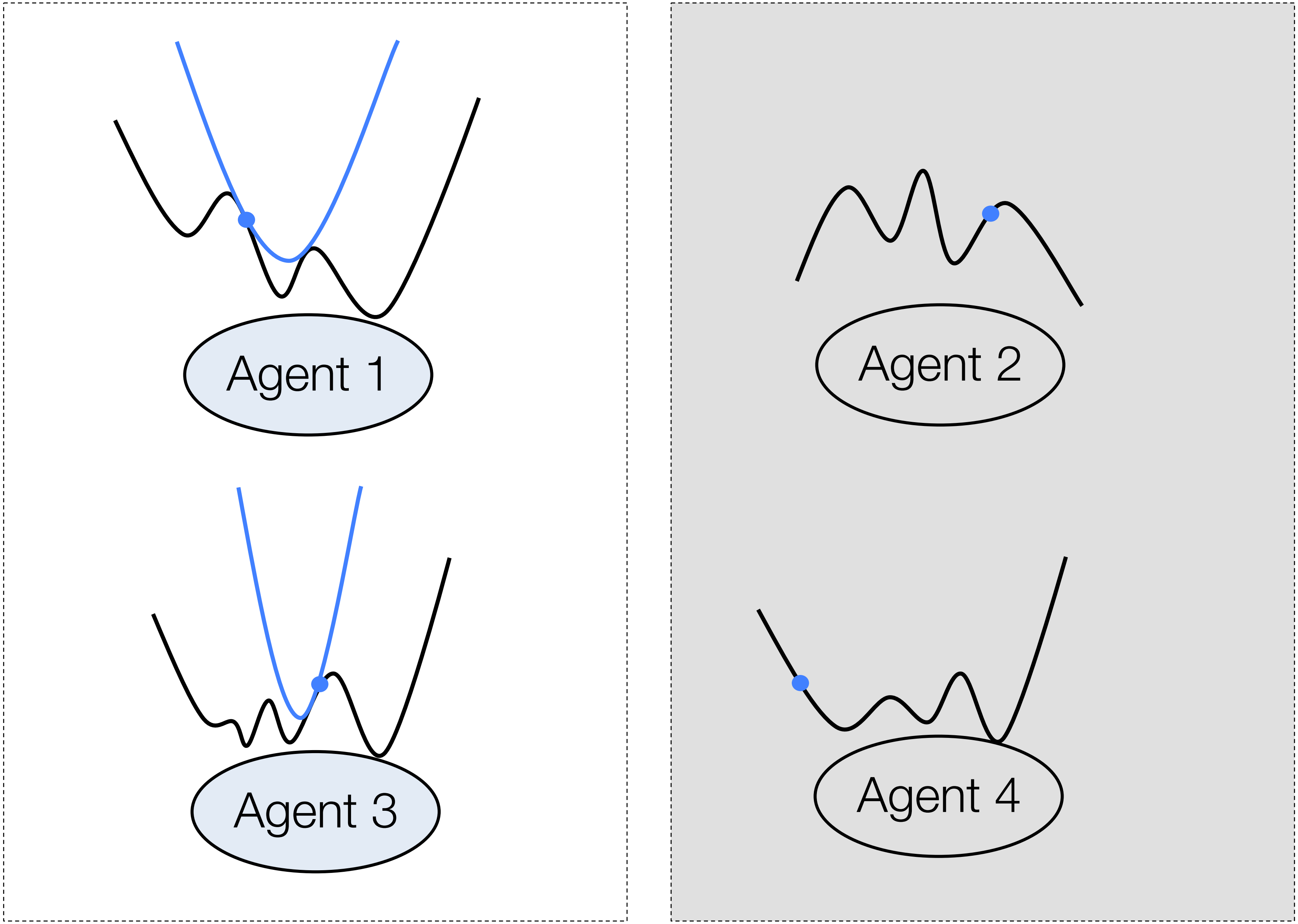
$$z_i^{l+1} = \operatorname{prox}_{\delta_{\mathcal{Z}_i}}^{c_i^l} \left(z_i^l - \frac{1}{c_i^l} \nabla_i L_{\rho^k}\left(z_1^{l+1}, \ldots, z_N^l, \mu^k\right) \right)$$

- Curvature c_i^l needs to be well-chosen
- Efficient for closed-form proxes (box, nonnegative orthant,…)
- Some degree of parallelisation depending on the coupling

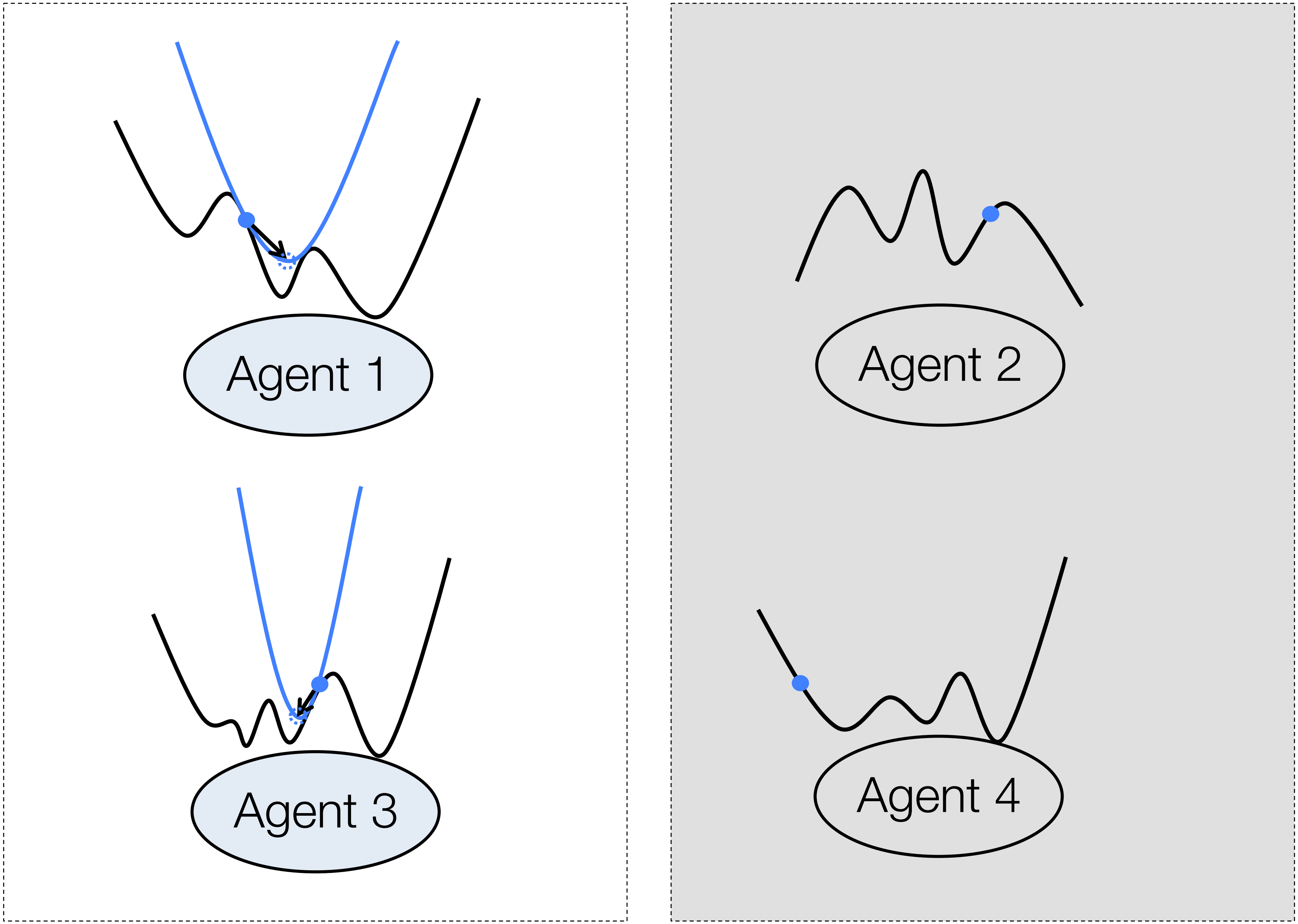
The Primal Loop



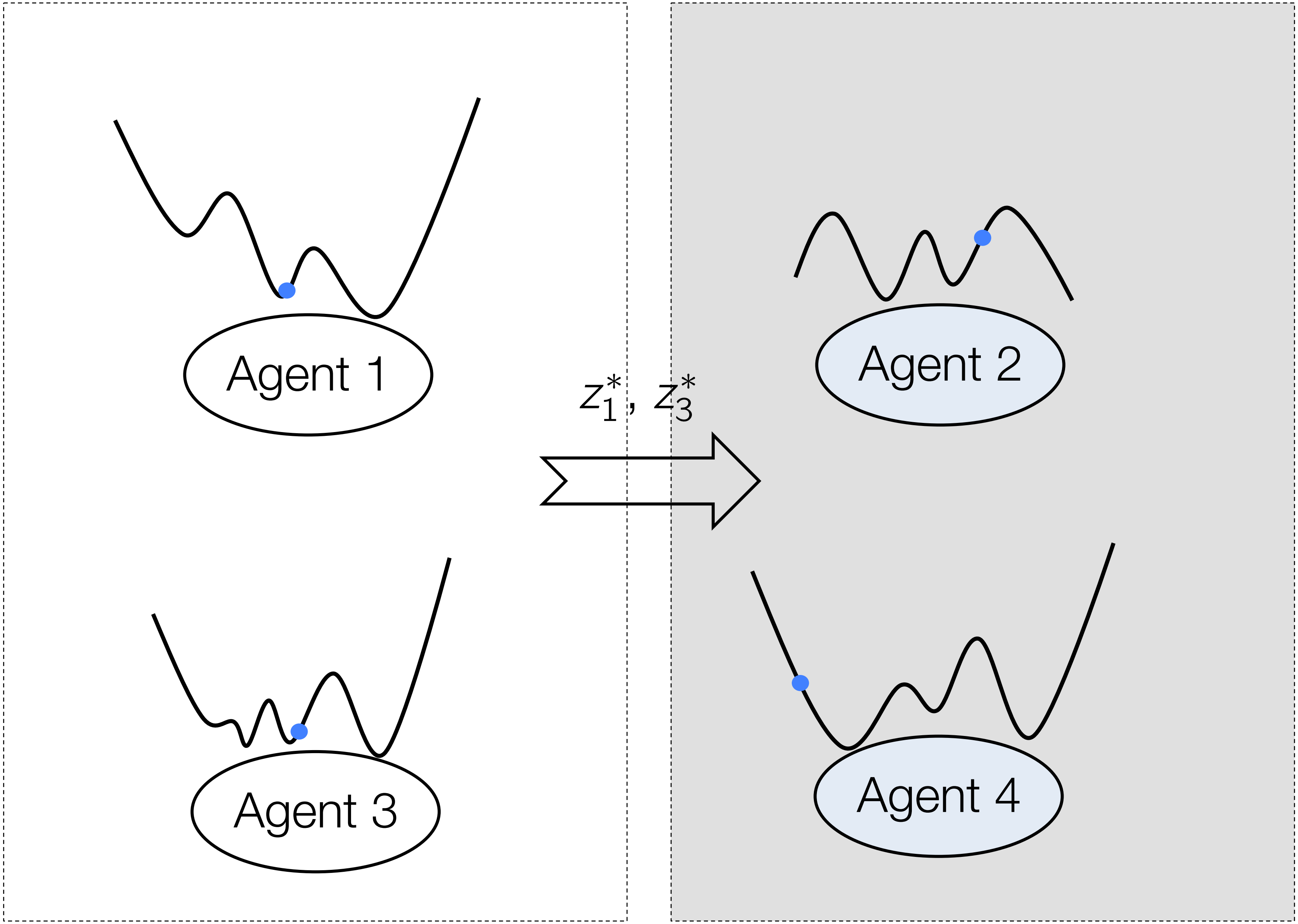
The Primal Loop



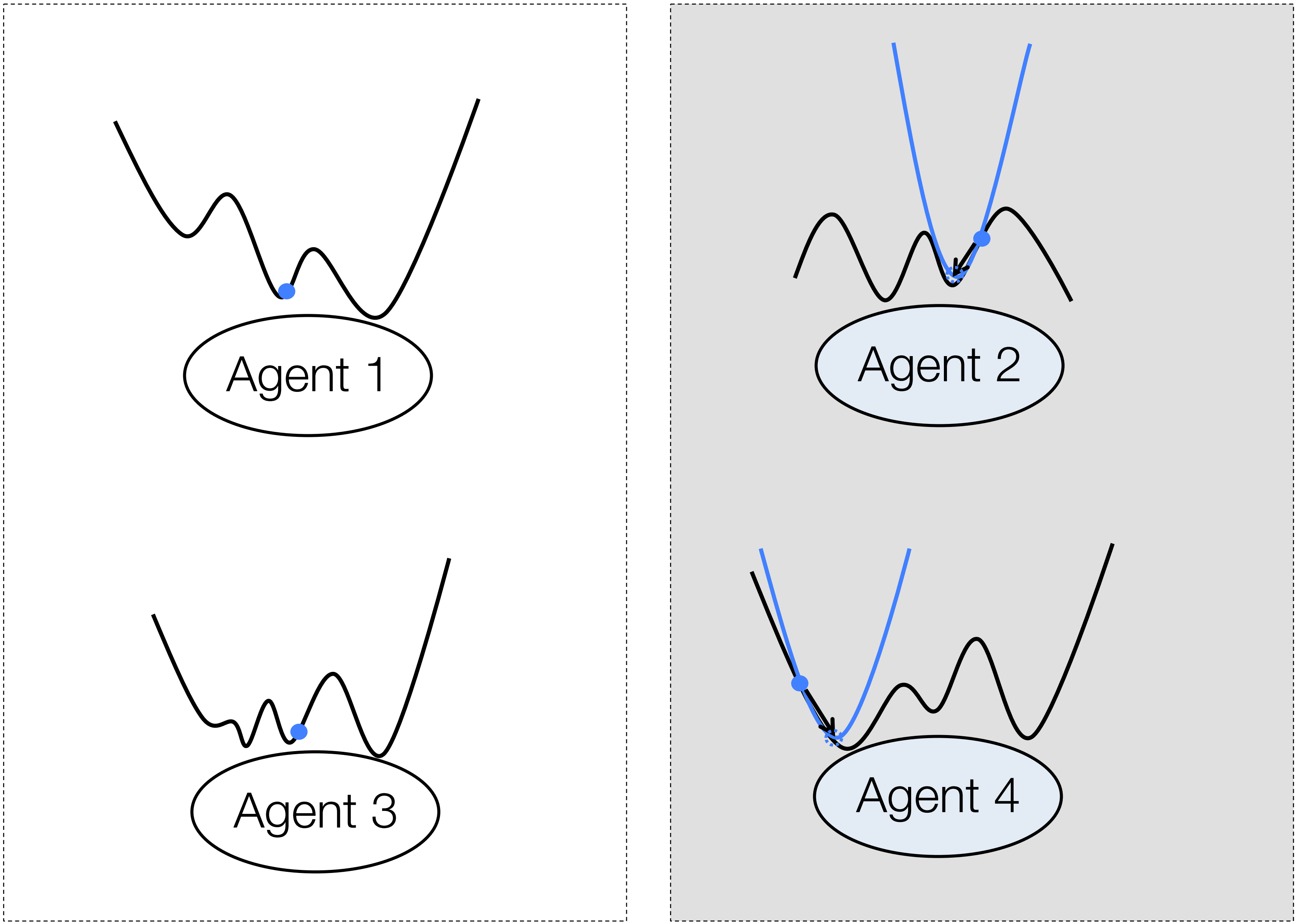
The Primal Loop



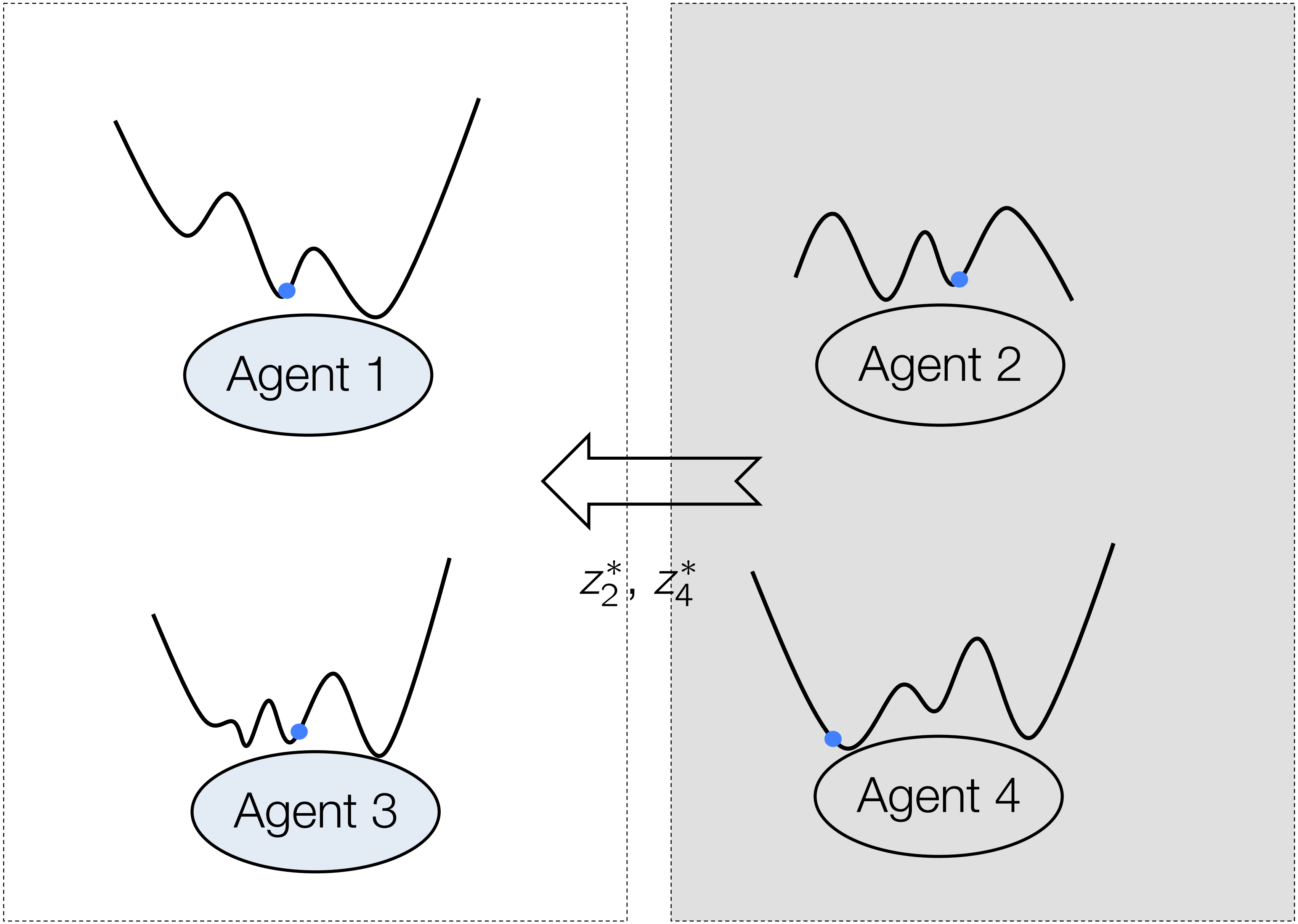
The Primal Loop



The Primal Loop



The Primal Loop



Summary

Proximal linearised alternations
(Parallelisable)

Dual update
(In parallel if cost coupling)

Penalty increase

Summary

Proximal linearised alternations
(Parallelisable)

Dual update
(In parallel if cost coupling)

Penalty increase

Summary

Proximal linearised alternations
(Parallelisable)

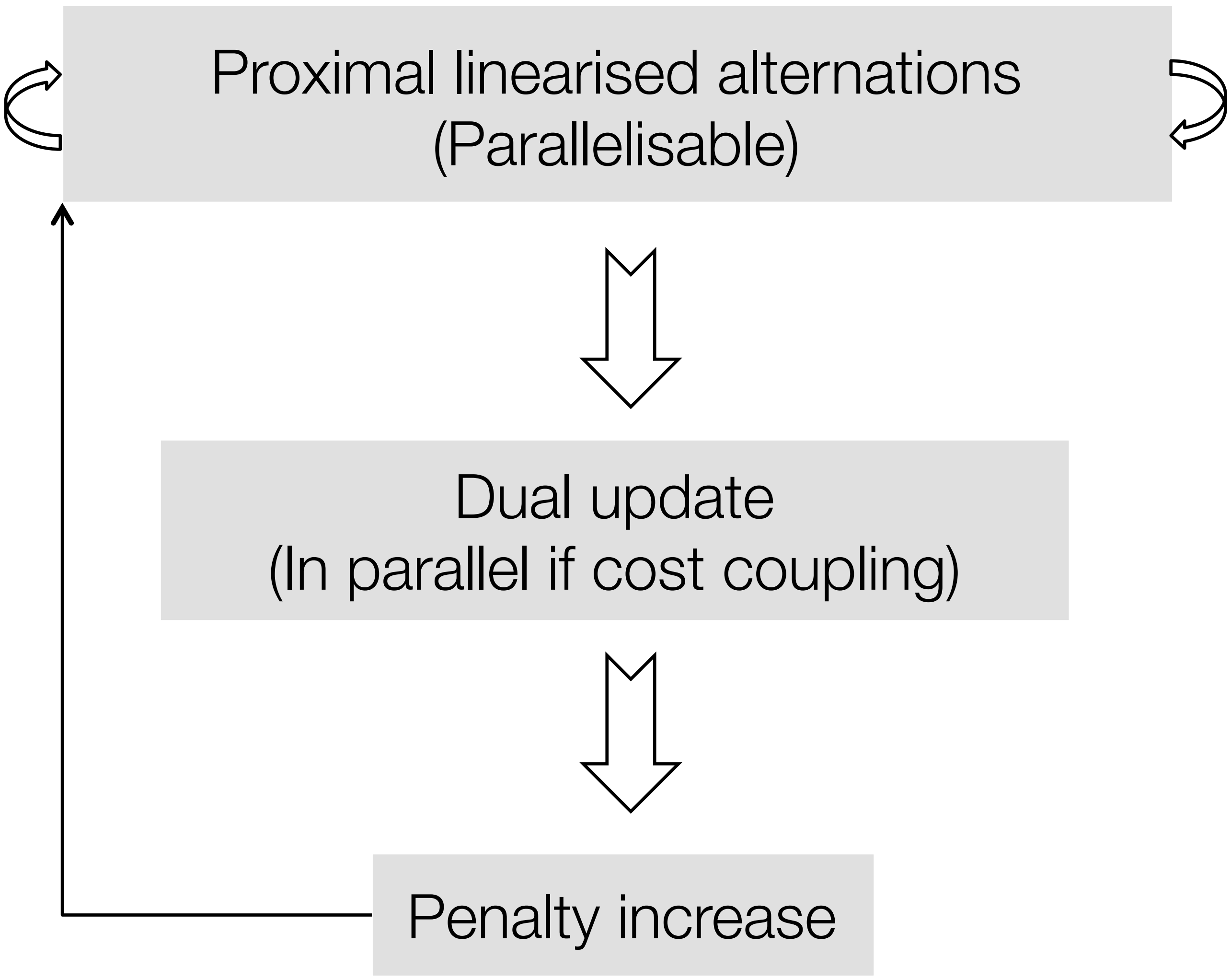
↓

Dual update
(In parallel if cost coupling)

↓

Penalty increase

Summary


Proximal linearised alternations
(Parallelisable)
Dual update
(In parallel if cost coupling)
Penalty increase

Motivation & Challenges

Algorithm Description

Convergence Analysis

Numerical Examples

A Recent Result about Descent Methods

L lower semi-continuous and **semi-algebraic**

Condition 1 : Sufficient decrease

$$L\left(z^{l+1}\right) + \beta \left\|z^{l+1} - z^{l}\right\|_2^2 \leq L\left(z^{l}\right)$$

Condition 2 : Relative error

$$\exists v^{l+1} \in \partial L(z^{l+1}),\ \left\|v^{l+1}\right\|_2 \leq \gamma \left\|z^{l+1} - z^{l}\right\|_2$$

(+ mild technical assumptions)

Thm: The sequence $\left\{z^{l}\right\}$ converges to a critical point z^* of L

[Attouch, Bolte, Svaiter, Math. Prog. 2013]

Monotonic Decrease in the Primal

Idea: Pick $c_i^l > \lambda_{\rho,\mu} + \beta_i$

$\lambda_{\rho,\mu}$ Lipshitz constant of $\nabla L_\rho(\cdot, \mu)$, β_i regularisation coefficient

Monotonic Decrease in the Primal

Idea: Pick $c_i^l > \lambda_{\rho,\mu} + \beta_i$

$\lambda_{\rho,\mu}$ Lipshitz constant of $\nabla L_\rho(\cdot, \mu)$, β_i regularisation coefficient

⇓ Descent Lemma

$$L_\rho\left(z_1^{l+1}, \ldots, z_i^{l+1}, \ldots, z_N^l, \mu\right) + \beta_i \left\|z_i^{l+1} - z_i^l\right\|_2^2 \leq L_\rho\left(z_1^{l+1}, \ldots, z_i^l, \ldots, z_N^l, \mu\right)$$

Monotonic Decrease in the Primal

Idea: Pick $c_i^l > \lambda_{\rho,\mu} + \beta_i$

$\lambda_{\rho,\mu}$ Lipshitz constant of $\nabla L_\rho(\cdot, \mu)$, β_i regularisation coefficient

$\Downarrow$ Descent Lemma

$$L_\rho\left(z_1^{l+1}, \ldots, z_i^{l+1}, \ldots, z_N^l, \mu\right) + \beta_i \left\|z_i^{l+1} - z_i^l\right\|_2^2 \leq L_\rho\left(z_1^{l+1}, \ldots, z_i^l, \ldots, z_N^l, \mu\right)$$

$\Downarrow$ Summation

$$L_\rho\left(z^{l+1}, \mu\right) + \beta \left\|z^{l+1} - z^l\right\|_2^2 \leq L_\rho\left(z^l, \mu\right)$$

In Practice: Local Backtracking Procedure

- Lipschitz constant:
  - may be difficult to compute
  - depends on the penalty parameter and the dual

Idea: Find smallest c_i^l such that

$$L_\rho\left(z_1^{l+1}, \ldots, z_i^{l+1}, \ldots, z_N\right) + \frac{\beta_i}{2}\left\|z_i^{l+1} - z_i^l\right\|_2^2$$

$$\leq L_\rho\left(z_1^{l+1}, \ldots, z_n^l\right) + \nabla_i L_\rho^\top\left(z_i^{l+1} - z_i^l\right) + \frac{c_i^l}{2}\left\|z_i^{l+1} - z_i^l\right\|_2^2$$

- Upper bound on local curvature (may be conservative)

- Efficient if proxes cheap to evaluate

Motivation & Challenges

Algorithm Description

Convergence Analysis

Numerical Examples

Example 1: Non-convex QP with quadratic constr.

Cost coupling

$$\text{minimise} \;\; J(z) := \sum_{i=1}^{N} z_i^\top H_i z_i + \underbrace{\sum_{i=1}^{N-1} z_i^\top G_{i,i+1} z_{i+1}}_{\text{Cost coupling}}$$

s.t.

$$z_i^\top z_i = \alpha_i$$

$$l_i \leq z_i \leq u_i, \;\; i \in \{1, \ldots, N\}$$

- Indefinite randomly generated hessians H_i and $G_{i,i+1}$
- **Test:** vary N and d, compare with IPOPT for same primal-dual initial guess

Example 1: Non-convex QP with quadratic constr.

$$\text{minimise } J(z) := \underbrace{\sum_{i=1}^{N} z_i^\top \underbrace{H_i}_{\text{Non-convex}} z_i}_{} + \overbrace{\sum_{i=1}^{N-1} z_i^\top \underbrace{G_{i,i+1}}_{\text{Non-convex}} z_{i+1}}^{\text{Cost coupling}}$$

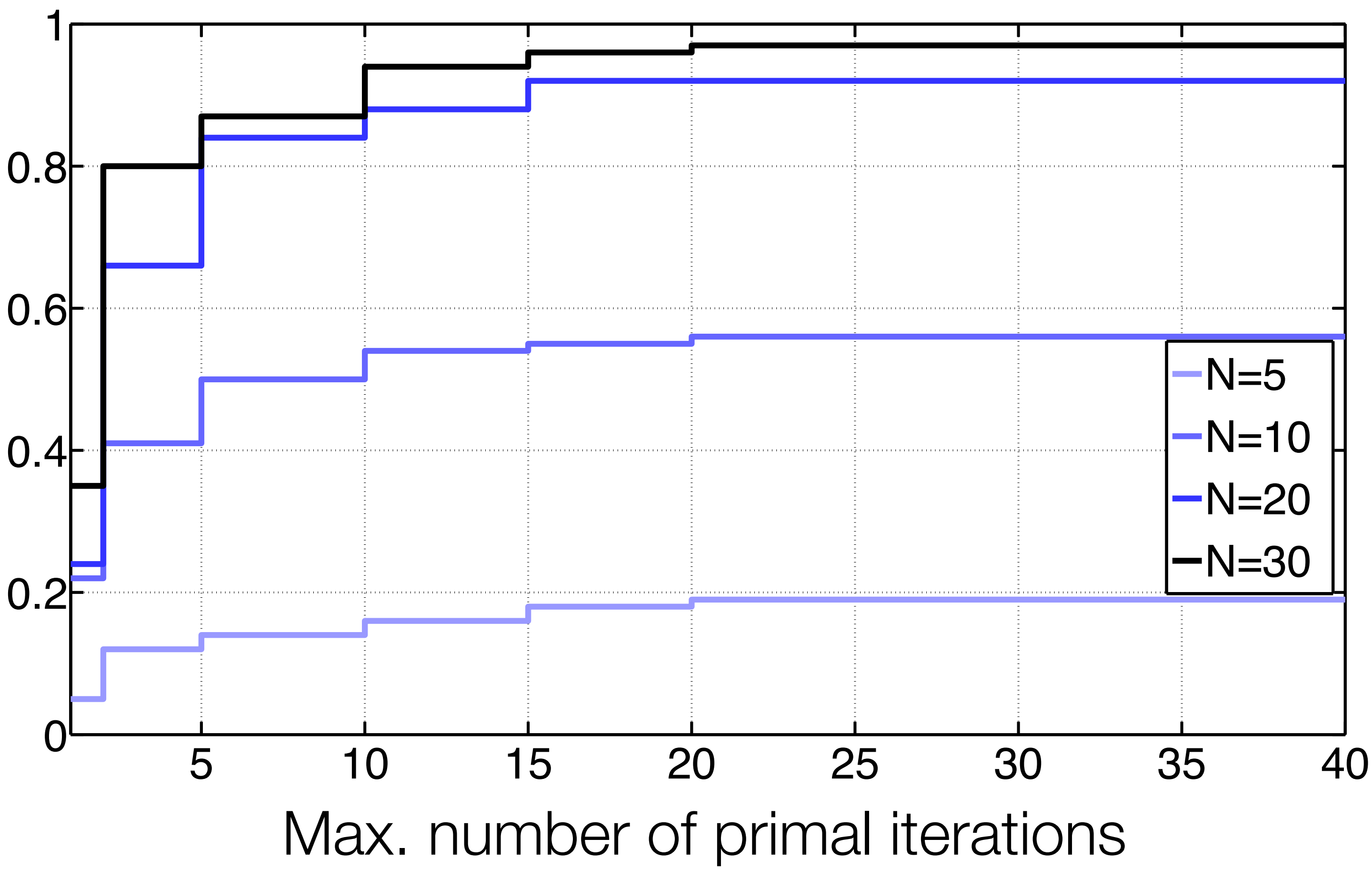
s.t.

$$\text{Non-convex} \begin{cases} z_i^\top z_i = \alpha_i \\ l_i \leq z_i \leq u_i, \; i \in \{1, \ldots, N\} \end{cases}$$

- Indefinite randomly generated hessians H_i and $G_{i,i+1}$
- **Test:** vary N and d, compare with IPOPT for same primal-dual initial guess

Example 1: Non-convex QP with quadratic constr.

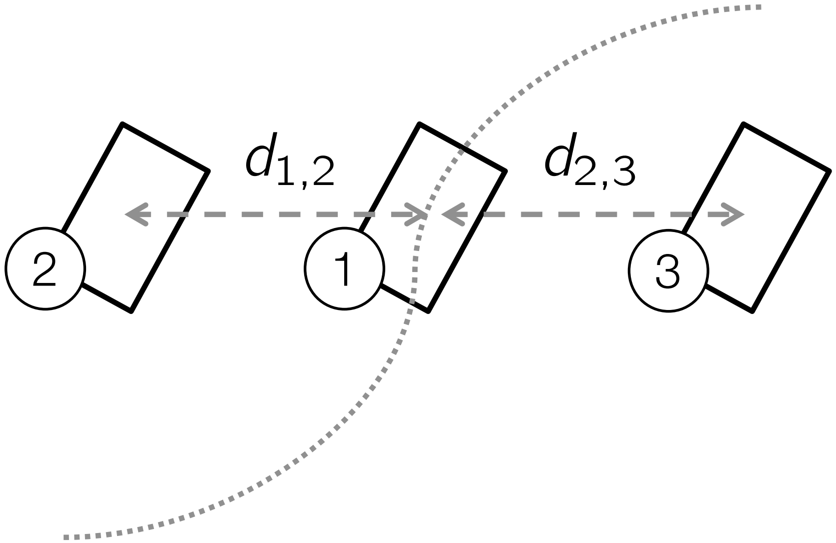
- Fix agent dimension d=2, vary number of agents N
- Performance criterion: $J(z^{\mathrm{AL}}) - J(z^{\mathrm{IP}}) < \theta_{\mathrm{pos}} < 0$



- But degrades as d increases…

Example 2: Collaborative tracking

- **Control objective:** track trajectory while staying in formation



Unicycle dynamics

$$\dot{x} = u_1 \cos\theta$$

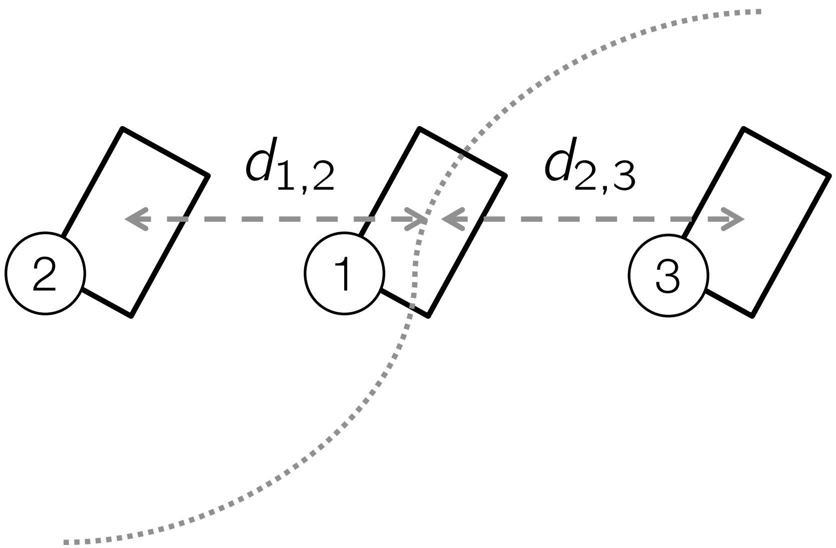
$$\dot{y} = u_1 \sin\theta$$

$$\dot{\theta} = u_2$$

$$\begin{aligned} \text{minimise} \;\; & \|z_1 - z_{\text{ref}}\| + \|z_2 - z_1 - d_{1,2}\| + \|z_3 - z_1 - d_{1,3}\| \\ \text{s.t.} \;\; & F(z_1) = 0, \; z_1 \in \mathcal{Z}_1 \\ & F(z_2) = 0, \; z_2 \in \mathcal{Z}_2 \\ & F(z_3) = 0, \; z_3 \in \mathcal{Z}_3 \end{aligned}$$

Example 2: Collaborative tracking

- **Control objective:** track trajectory while staying in formation



Unicycle dynamics

$$\dot{x} = u_1 \cos\theta$$

$$\dot{y} = u_1 \sin\theta$$

$$\dot{\theta} = u_2$$

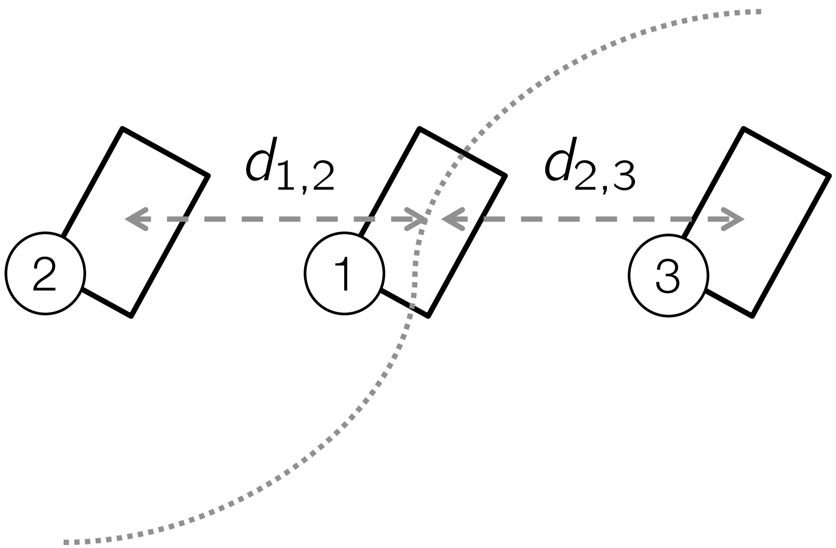
Path-following — Formation keeping

$$\text{minimise } \underbrace{\|z_1 - z_{\text{ref}}\|}_{\text{Path-following}} + \underbrace{\|z_2 - z_1 - d_{1,2}\| + \|z_3 - z_1 - d_{1,3}\|}_{\text{Formation keeping}}$$

$$\left.\begin{aligned} \text{s.t. } F(z_1) &= 0,\ z_1 \in \mathcal{Z}_1 \\ F(z_2) &= 0,\ z_2 \in \mathcal{Z}_2 \\ F(z_3) &= 0,\ z_3 \in \mathcal{Z}_3 \end{aligned}\right\} \text{Decoupled dynamics}$$

Example 2: Collaborative tracking

- **Control objective:** track trajectory while staying in formation



Unicycle dynamics

$$\dot{x} = u_1 \cos\theta$$
$$\dot{y} = u_1 \sin\theta$$
$$\dot{\theta} = u_2$$

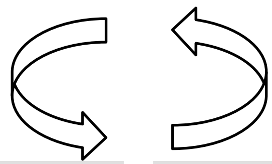
$$\begin{aligned} \text{minimise} \quad & \|z_1 - z_{\text{ref}}\| + \|z_2 - z_1 - d_{1,2}\| + \|z_3 - z_1 - d_{1,3}\| \\ \text{s.t.} \quad & F(z_1) = 0,\ z_1 \in \mathcal{Z}_1 \\ & F(z_2) = 0,\ z_2 \in \mathcal{Z}_2 \\ & F(z_3) = 0,\ z_3 \in \mathcal{Z}_3 \end{aligned}$$

Fix z_1, decomposes in z_2, z_3

Example 2: Collaborative tracking

$$\underset{z_1, z_2, z_3}{\text{minimise}} \; L_\rho(z_1, z_2, z_3, \mu_1, \mu_2, \mu_3) + \delta_{\mathcal{Z}_1}(z_1) + \delta_{\mathcal{Z}_2}(z_2) + \delta_{\mathcal{Z}_3}(z_3)$$

$$z_1^{k+1} \leftarrow \text{prox}_{\delta_{\mathcal{Z}_1}} \left(z_1^k - \frac{1}{c_1^k} \nabla L_\rho \left(z_1^k, z_2^k, z_3^k \right) \right)$$



$$z_2^{k+1} \leftarrow \text{prox}_{\delta_{\mathcal{Z}_2}} \left(z_2^k - \frac{1}{c_2^k} \nabla L_\rho \left(z_1^{k+1}, z_2^k, z_3^k \right) \right)$$

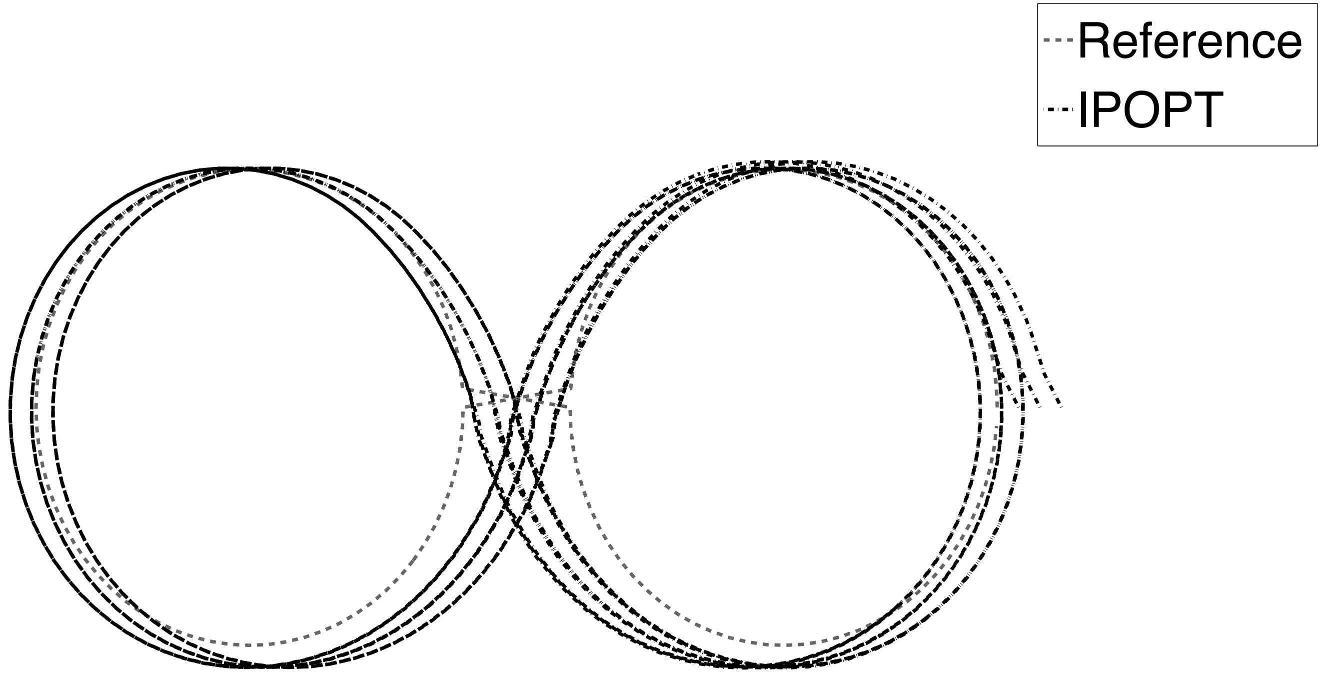
$$z_3^{k+1} \leftarrow \text{prox}_{\delta_{\mathcal{Z}_3}} \left(z_3^k - \frac{1}{c_3^k} \nabla L_\rho \left(z_1^{k+1}, z_2^k, z_3^k \right) \right)$$

Update μ_1

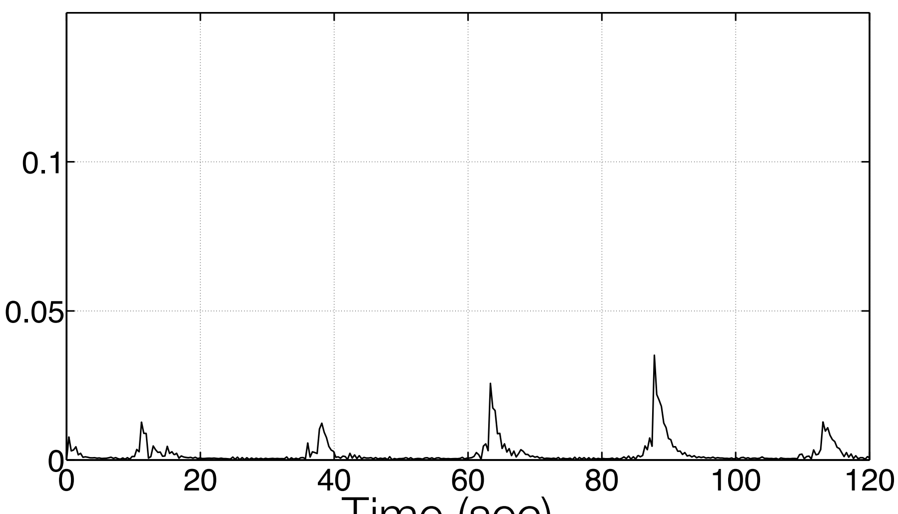
Update μ_2

Update μ_3

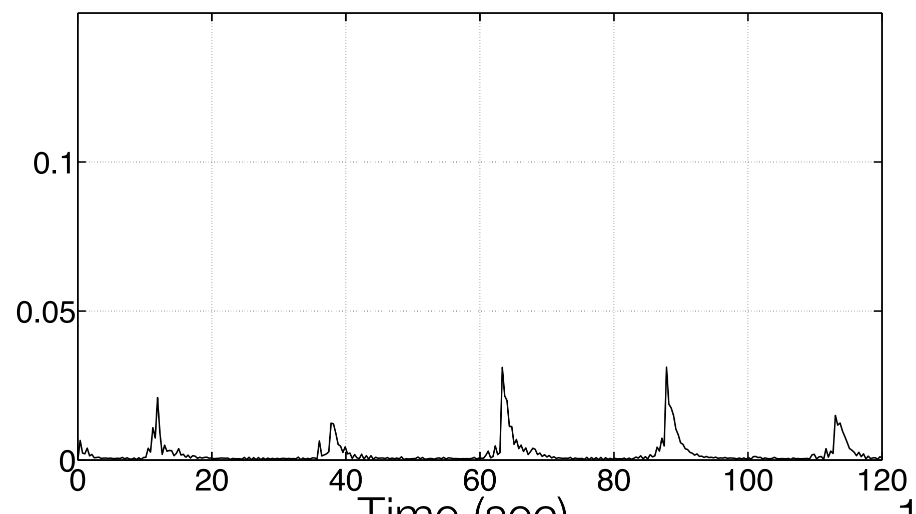
Example 2: Collaborative tracking



$\|z_2 - z_1 - d_{1,2}\|$

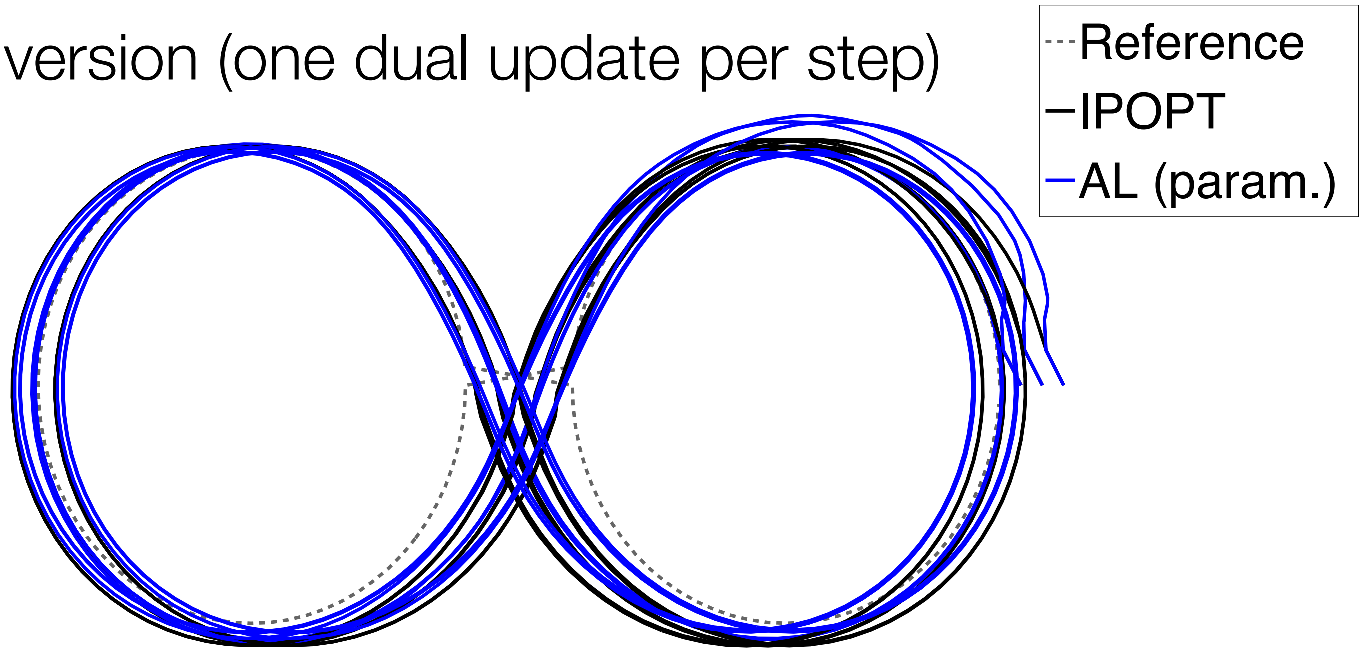


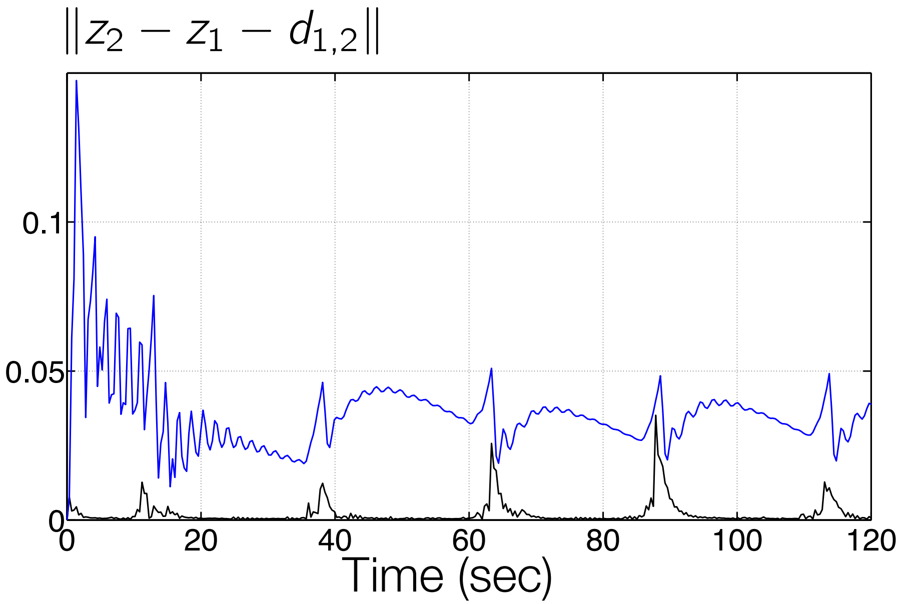
$\|z_3 - z_1 - d_{1,3}\|$

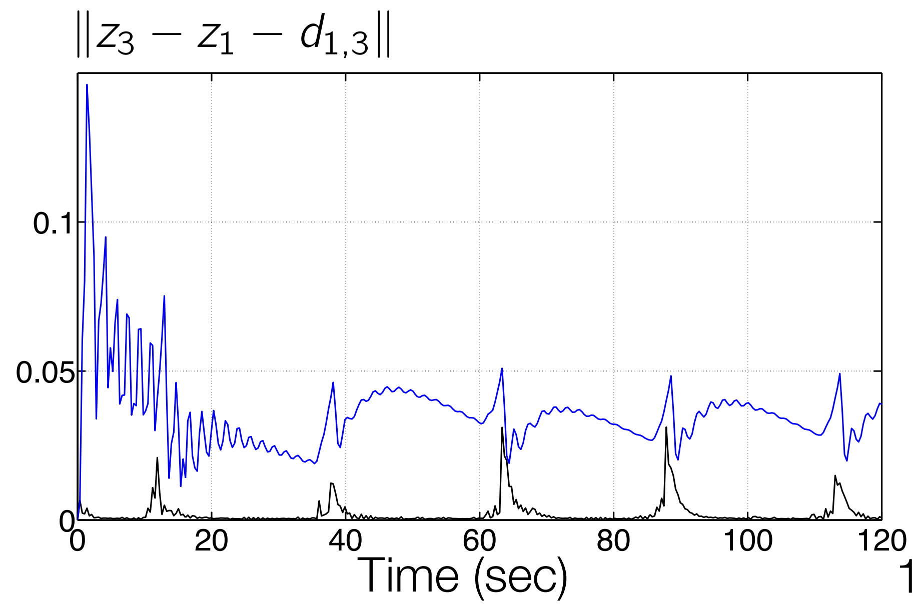


Example 2: Collaborative tracking

- Suboptimal version (one dual update per step)







Conclusion

- Distributed non-convex optimisation within the augmented Lagrangian framework
- **Key idea:** Address primal decomposition with proximal linearised alternations
- In practice, local backtracking procedure
- Good performance on very sparse non-convex programs (many small agents)
- **Future work:**
    - Extension to the parametric case (NMPC,…)
    - Stopping criterion for inner loop