Decentralized non-convex optimization via bi-level SQP and ADMM

Alternating Direction Method of Multipliers Session, IEEE CDC 2022

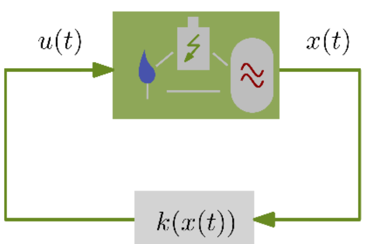
Gösta Stomberg, Alexander Engelmann, Timm Faulwasser

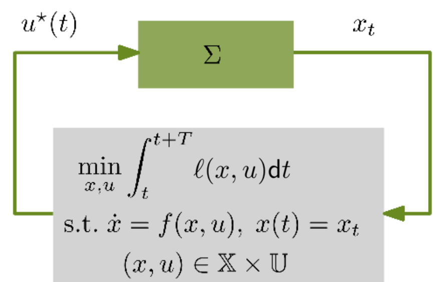
Institute for Energy Systems, Energy Efficiency and Energy Economics

TU Dortmund University

goesta.stomberg@tu-dortmund.de

TuAT08.1

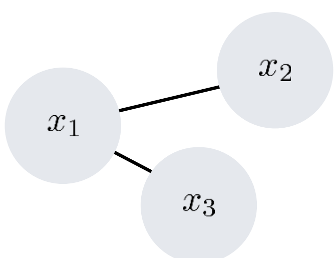




Problem Statement

Partially separable NLP

$$
\begin{array}{llll}
\min\limits_{x_1,\dots,x_S} & \sum\limits_{i\in\mathcal{S}} f_i(x_i) & & f_i \in \mathcal{C}^2 : \mathbb{R}^{n_i} \to \mathbb{R} \\
\text{s.t.} & \cancel{g_i(x_i) = 0 \quad | \, \nu_i \quad \forall i \in \mathcal{S} \quad g_i \in \mathcal{C}^2 : \mathbb{R}^{n_i} \to \mathbb{R}^{n_{gi}}} & & g_i(x_i) \text{ in paper} \\
 & h_i(x_i) \le 0 \quad | \, \mu_i & \forall i \in \mathcal{S} & h_i \in \mathcal{C}^2 : \mathbb{R}^{n_i} \to \mathbb{R}^{n_{hi}} \\
 & \sum\limits_{i\in\mathcal{S}} E_i x_i = c \quad | \, \lambda & & E_i \in \mathbb{R}^{n_c \times n_i}
\end{array}
$$



Desirable algorithmic properties

- Communication only between neighbors (decentralized method)
- Convergence guarantees for non-convex problems
- Low-complexity computations per subsystem

How to design a suitable method?

Bertsekas, D. and Tsitsiklis, J. “Parallel and distributed computation: numerical methods.” Athena Scientific, 2015

Sequential Quadratic Programming (SQP)

Centralized NLP

$$\begin{aligned} \min_{x} \quad & f(x) \\ \text{s.t.} \quad & h(x) \leq 0 \mid \mu \\ & Ex = c \mid \lambda \end{aligned}$$

$$p^k = \begin{bmatrix} x^k \\ \mu^k \\ \lambda^k \end{bmatrix}$$

QP approximation (convex)

$$H^k \approx \nabla^2_{xx} L(p^k)$$

$$\begin{aligned} \min_{s} \quad & \frac{1}{2} s^\top H^k s + \nabla f^{k\top} s \\ \text{s.t.} \quad & h^k + \nabla h^{k\top} s \leq 0 \mid \mu^{QP} \\ & E(x^k + s) = c \mid \lambda^{QP} \end{aligned}$$

$$d^k = \begin{bmatrix} s^k \\ \mu^{QP,k} - \mu^k \\ \lambda^{QP,k} - \lambda^k \end{bmatrix}$$

SQP

- Evaluate sensitivites
- Solve QP for d^k
- $p^{k+1} = p^k + d^k$

Local convergence via Newton's method

Boggs, P. and Tolle, J. “Sequential quadratic programming.” *Acta numerica,* 1995
Nocedal, J. and Wright, S. “Numerical Optimization.” *Springer Science & Business Media*, 2006

Sequential Quadratic Programming (SQP)

KKT conditions

$$\nabla_x L(p) = 0$$
$$h(x) \leq 0,\ \mu \geq 0,\ \mu^\top h(x) = 0$$
$$Ex - c = 0$$

$$F(p) \doteq \begin{bmatrix} \nabla_x L(p) \\ \mathrm{min}(-h(x), \mu) \\ Ex - c \end{bmatrix} = 0$$

Newton: $p^{k+1} = p^k - \nabla(F^k)^{-1} F^k$

Assumption 1 *(KKT point)*
The KKT point $p^\star$ satisfies

i) Strict complementarity

ii) SOSC

iii) LICQ

Local convergence of SQP

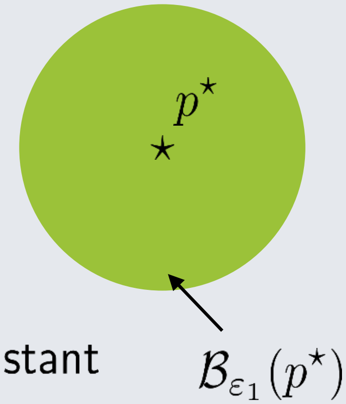
If $p^0 \in \mathcal{B}_{\varepsilon_1}(p^\star)$, then:

i) $\nabla F^{k^{-1}}$ exists

ii) $p^{k+1} \in \mathcal{B}_{\varepsilon_1}(p^\star)$

iii) Active set $\mathcal{A}^k$ stays constant

iv) $\{p^k\} \to p^\star$ q-superlinearly



Apply SQP to partially separable NLP?

Boggs, P. and Tolle, J. “Sequential quadratic programming.” *Acta numerica,* 1995
Nocedal, J. and Wright, S. “Numerical Optimization.” *Springer Science & Business Media*, 2006

Decentralized SQP

Partially separable NLP

$$\min_{x_1,\dots,x_S} \sum_{i\in\mathcal{S}} f_i(x_i)$$

$$\text{s.t.} \quad h_i(x_i) \le 0 \mid \mu_i$$

$$\sum_{i\in\mathcal{S}} E_i x_i = c \mid \lambda$$

$$p^k = \begin{bmatrix} x^k \\ \mu^k \\ \lambda^k \end{bmatrix}$$

QP approximation (convex)

$$H_i^k \approx \nabla^2_{x_i x_i} L(p^k)$$

$$\min_{s_1,\dots,s_S} \sum_{i\in\mathcal{S}} \frac{1}{2} s_i^\top H_i^k s_i + \nabla f_i^{k\top} s_i$$

$$\text{s.t.} \quad h_i^k + \nabla h_i^{k\top} s_i \le 0 \mid \mu_i^{QP}$$

$$\sum_{i\in\mathcal{S}} E_i(x_i^k + s_i) = c \mid \lambda^{QP}$$

Decentralized SQP (d-SQP)

Evaluate sensitivities per subsystem

Solve partially separable convex QP

$$x_i^{k+1} = x_i^k + s_i \qquad \mu_i^{k+1} = \mu_i^{QP} \qquad \lambda^{k+1} = \lambda^{QP}$$

Solve in decentralized fashion?

Solve the QP with ADMM to obtain a decentralized SQP scheme

Alternating Direction Method of Multipliers (ADMM)

technische universität dortmund

Inner QP

$$\min_{s_1,\dots,s_S} \sum_{i\in\mathcal{S}} \frac{1}{2}{s_i}^\top H_i^k s_i + \nabla f_i^{k\top} s_i$$

$$\text{s.t.} \quad \left. h_i^k + \nabla h_i^{k\top} s_i \le 0 \right\} \mathbb{S}_i$$

$$\left. \sum_{i\in\mathcal{S}} E_i(x_i^k + s_i) = c \right\} \mathbb{E}$$

$$\bar{s}_i \in \mathbb{R}^{ni}$$

$$\min_{s_i,\bar{s}_i,i\in\mathcal{S}} \sum_{i\in\mathcal{S}} \frac{1}{2}{s_i}^\top H_i^k s_i + \nabla f_i^{k\top} s_i$$

$$\text{s.t.} \quad s_i \in \mathbb{S}_i$$

$$s_i - \bar{s}_i = 0 \mid \gamma_i$$

$$\sum_{i\in\mathcal{S}} E_i(x_i^k + \bar{s}_i) = c$$

$$L_\rho(s,\bar{s},\nu) = \sum_{i\in\mathcal{S}} L_{\rho,i}(s_i,\bar{s}_i,\nu_i) = \sum_{i\in\mathcal{S}} \frac{1}{2}{s_i}^\top H_i^k s_i + \nabla f_i^{k\top} s_i + \gamma_i^\top(s_i - \bar{s}_i) + \frac{\rho}{2}\|s_i - \bar{s}_i\|_2^2$$

ADMM

$$s_i^{l+1} = \operatorname*{argmin}_{s_i\in\mathbb{S}_i} L_{\rho,i}(s_i, \bar{s}_i^l, \gamma_i^l)$$

$$\bar{s}^{l+1} = \operatorname*{argmin}_{\bar{s}\in\mathbb{E}} \sum_{i\in\mathcal{S}} L_{\rho,i}(s_i^{l+1}, \bar{s}_i, \gamma_i^l)$$

$$\gamma_i^{l+1} = \gamma_i^l + \rho(s_i^{l+1} - \bar{s}_i^{l+1})$$

Decentralized averaging

Convex inner QP $\implies$ convergence

Boyd, S., Parikh, N., Chu, E., Peleato, B. and Eckstein, J. “Distributed optimization and statistical learning via the alternating direction method of multipliers.” *Found. Trends Mach. Learn.,* 2011

Decentralized SQP

technische universität dortmund

Partially separable NLP

$$\min_{x_1,\dots,x_S} \sum_{i\in\mathcal{S}} f_i(x_i)$$
$$\text{s.t.} \quad h_i(x_i) \le 0 \mid \mu_i$$
$$\sum_{i\in\mathcal{S}} E_i x_i = c \mid \lambda$$

$$p^k = \begin{bmatrix} x^k \\ \mu^k \\ \lambda^k \end{bmatrix}$$

QP approximation (convex)

$$H_i^k \approx \nabla^2_{x_i x_i} L(p^k)$$

$$\min_{s_1,\dots,s_S} \sum_{i\in\mathcal{S}} \frac{1}{2} {s_i}^\top H_i^k s_i + \nabla f_i^{k\top} s_i$$
$$\text{s.t.} \quad h_i^k + \nabla h_i^{k\top} s_i \le 0 \mid \mu_i^{QP}$$
$$\sum_{i\in\mathcal{S}} E_i(x_i^k + s_i) = c \mid \lambda^{QP}$$

Decentralized SQP (d-SQP)

Evaluate sensitivities per subsystem

Solve partially separable convex QP with ADMM

Terminate early?

$s_i, \bar{s}_i, \lambda$

$$x_i^{k+1} = x_i^k + \bar{s}_i \qquad \mu_i^{k+1} = \mu_i^{QP} \qquad \lambda^{k+1} = \lambda^{QP}$$

Inexact decentralized SQP steps! Convergence?

Inexact SQP

technische universität
dortmund

KKT

$$F(p) = \begin{bmatrix} \nabla_x L(p) \\ \min(-h(x), \mu) \\ Ex - c \end{bmatrix}$$

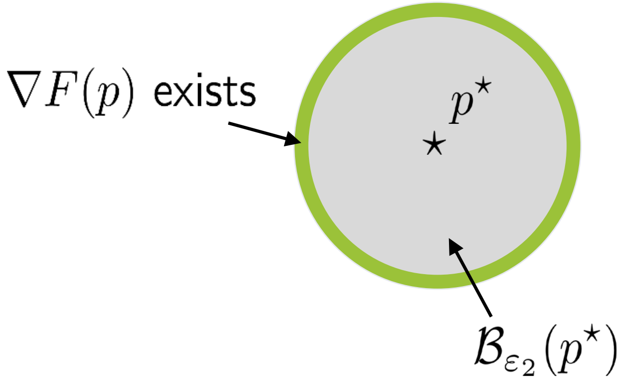
Newton: $F^k + \nabla F^k d^k = 0$

Inexact Newton: $\|F^k + \nabla F^k d^k\| \leq \eta^k \|F^k\|$ (IN)

Lemma *(Inexact SQP convergence)*
Let Assumption 1 hold and let d^k satisfy (IN). If $p^0 \in \mathcal{B}_{\varepsilon_2}(p^\star)$, then:

i) $\{p^k\} \to p^\star$ *q-linearly, if* $\eta^k \leq \eta$

ii) $\{p^k\} \to p^\star$ *q-superlinearly, if* $\eta^k \to 0$



How to evaluate (IN) outside $\mathcal{B}_{\varepsilon_1}$?

Dembo, R. S., Eisenstat, S. C. and Steihaug, T. “Inexact newton methods" *SIAM J. Nume Anal.,* 1982
Morini, B. “Convergence behavior of inexact newton methods.” *Math. Comput.,* 1999

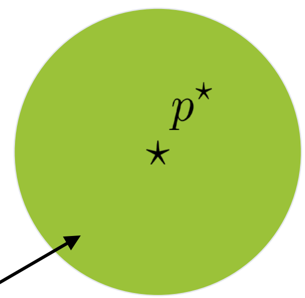
Evaluating the Stopping Criterion

KKT

$$F(p) = \begin{bmatrix} \nabla_x L(p) \\ \min(-h(x), \mu) \\ Ex - c \end{bmatrix}$$

Newton: $F^k + \nabla F^k d^k = 0$

Inexact Newton: $\|F^k + \nabla F^k d^k\| \leq \eta^k \|F^k\|$ (IN)



$\nabla_p \min(-h(x), \mu) = ?$

Optimal active set $\mathcal{A}^\star \implies \min(-h(x), \mu) = 0$

Modified stopping criterion (decentralized)

$$\tilde{F}(p) \doteq \begin{bmatrix} \nabla_x L(p) \\ Ex - c \end{bmatrix}$$

$\|\tilde{F}^k + \nabla \tilde{F}^k d^k\| \leq \eta^k \|\tilde{F}^k\|$ (SC)

If ADMM terminates at $\mathcal{A}^\star$, then (SC) $\implies$ (IN)

Decentralized SQP

Decentralized SQP (d-SQP)

Evaluate sensitivities per subsystem

Solve partially separable convex QP with ADMM until $\|\tilde{F}^k + \nabla \tilde{F}^k d^k\| \leq \eta^k \|\tilde{F}^k\|$

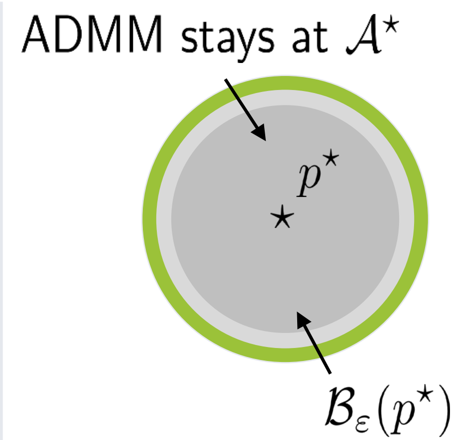
$s_i, \bar{s}_i, \lambda$

$x_i^{k+1} = x_i^k + \bar{s}_i \quad \mu_i^{k+1} = \mu_i^{QP} \quad \lambda^{k+1} = \lambda^{QP}$

Theorem *(d-SQP convergence)*
Let Assumption 1 hold and initialize ADMM with $s_i = 0$ *and* $\gamma_i = E_i^\top \lambda^k$*. If* $p^0 \in \mathcal{B}_\varepsilon(p^\star)$*, then:*

i) $\{p^k\} \to p^\star$ *q-linearly in the outer iterations, if* $\eta^k \leq \eta$

ii) $\{p^k\} \to p^\star$ *q-superlinearly in the outer iterations, if* $\eta^k \to 0$



d-SQP is guaranteed to converge locally for non-convex problems!

Numerical Example: AC-OPF

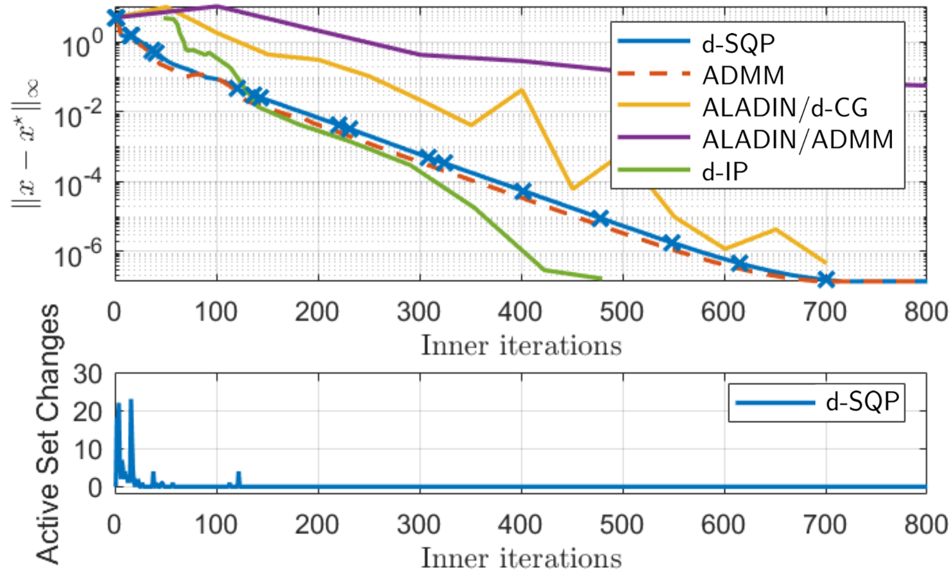
Partially separable NLP

$$\min_{x_1,\dots,x_S} \sum_{i\in\mathcal{S}} f_i(x_i) \qquad x \in \mathbb{R}^{576}$$

$$\text{s.t.} \quad g_i(x_i) = 0 \qquad g(x) \in \mathbb{R}^{470}$$

$$h_i(x_i) \le 0 \qquad h(x) \in \mathbb{R}^{792}$$

$$\sum_{i\in\mathcal{S}} E_i x_i = 0 \qquad E \in \mathbb{R}^{52\times 576}$$



Method	Only neighbor-to-neighbor communication	Proven convergence
d-SQP	yes	yes
ADMM	yes	no
ALADIN/d-CG	no	yes
ALADIN/ADMM	yes	yes
d-IP	no	yes

Computation time until $\|x - x^\star\| < 10^{-6}$
Matlab, one computer for all subsystems

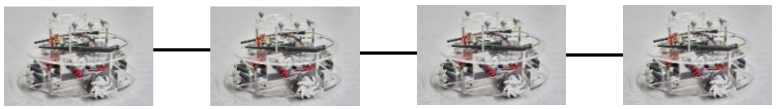
d-SQP	30 s
ADMM	67 s

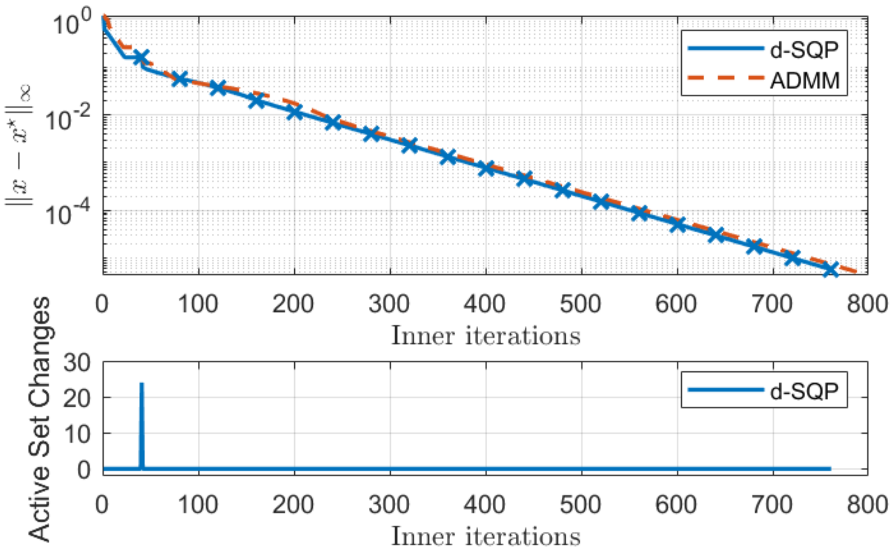
d-SQP shows competitive performance while only solving QPs

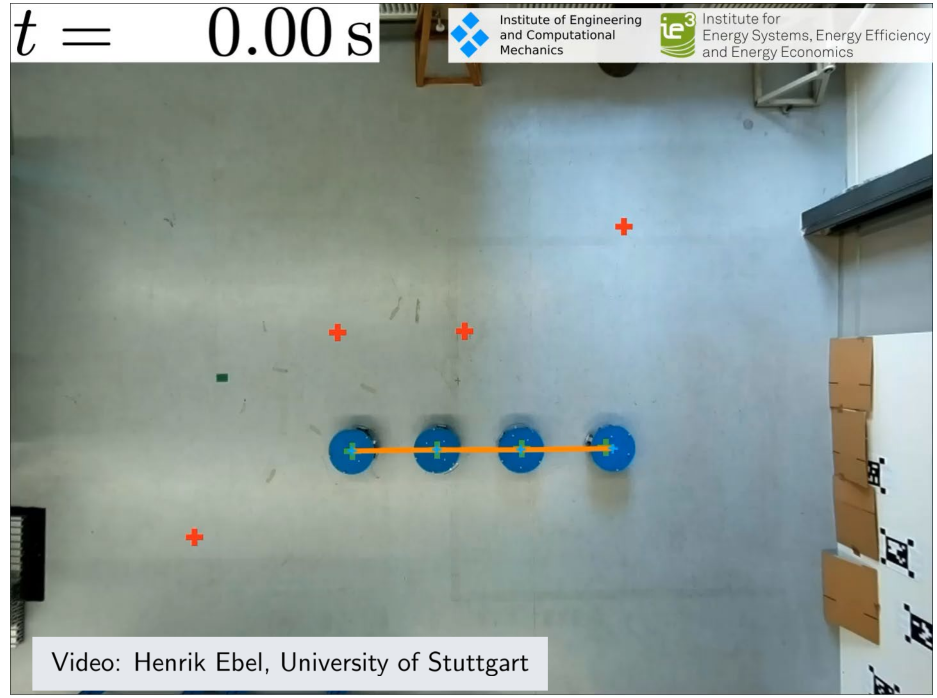
Erseghe, T. "Distributed optimal power flow using ADMM." IEEE Trans. Power Syst., 2014
Frank, S and Rebennack, S. "An introduction to optimal power flow: theory, formulation, and examples." *IIE transactions*, 2016
Engelmann, A. et al. "Decomposition of nonconvex optimization via bi-level distributed ALADIN." *IEEE Trans. Control Netw. Syst.*, 2020

Application Example: Distributed MPC

Mobile robots (4 subsystems)







Computation time to solve optimal control problem
C++, one computer per robot

median	33.81 ms
maximum	53.06 ms

$< \Delta t = 200\,\mathrm{ms}!$

Promising results for distributed NMPC

Stomberg, G., Ebel, H., Faulwasser, T. and Eberhard, P. "Distributed Model Predictive Formation Control in Real-Time", arXiv, 2022

Summary

Decentralized SQP (d-SQP)

Evaluate sensitivities per subsystem

Solve partially separable convex QP with ADMM

$s_i, \bar{s}_i, \lambda$

$x_i^{k+1} = x_i^k + \bar{s}_i \quad \mu_i^{k+1} = \mu_i^{QP} \quad \lambda^{k+1} = \lambda^{QP}$

Key features and outlook

- Local convergence gurantees for non-convex NLPs
- Communication only between coupled subsystems (decentralized method)
- Solves convex QPs on a subsystem level
- Open problem: globalization

Thank you

Stomberg, G., Engelmann, A. and Faulwasser T. “Decentralized non-convex optimization via bi-level SQP and ADMM”, 61st CDC, 2022
Stomberg, G., Ebel, H., Faulwasser, T. and Eberhard, P. “Distributed Model Predictive Formation Control in Real-Time”, arXiv, 2022

References

Dembo, R. S., Eisenstat, S. C. and Steihaug, T. “Inexact newton methods” *SIAM J. Nume Anal.,* vol. 10, no. 2, pp. 400-408, 1982

Boggs, P. and Tolle, J. “Sequential quadratic programming.” *Acta numerica,* vol. 4, pp. 1-51, 1995

Morini, B. “Convergence behavior of inexact newton methods.” *Math. Comput.,* vol. 68, no. 228, pp. 1604-1613, 1999

Nocedal, J. and Wright, S. “Numerical Optimization.” *Springer Science & Business Media*, 2006

Boyd, S., Parikh, N., Chu, E., Peleato, B. and Eckstein, J. “Distributed optimization and statistical learning via the alternating direction method of multipliers.” *Found. Trends Mach. Learn.,* vol. 3, no. 1, pp. 1–122, 2011

Erseghe, T. “Distributed optimal power flow using ADMM.” IEEE Trans. Power Syst., vol. 29, no. 5, pp. 2370-2380, 2014

Frank, S and Rebennack, S. "An introduction to optimal power flow: Theory, formulation, and examples." *IIE transactions* 48.12: 1172-1197, 2016

Engelmann, A., Jiang, Y., Houska, B. and Faulwasser, T. “Decomposition of nonconvex optimization via bi-level distributed ALADIN.” *IEEE Trans. Control Netw. Syst.,* vol. 7, no. 4, pp. 1848–1858, 2020

Stomberg, G., Ebel, H., Faulwasser, T. and Eberhard, P. “Distributed Model Predictive Formation Control in Real-Time”, arXiv, 2022