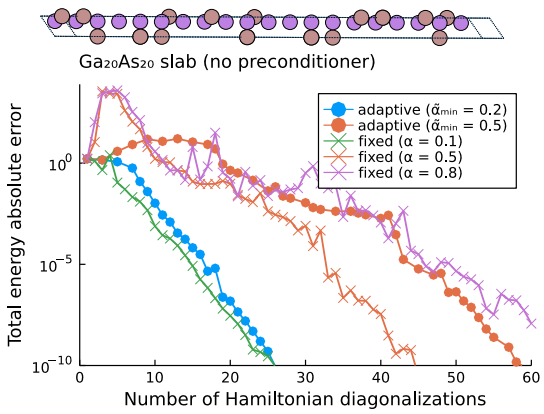
A guaranteed and efficient line search for self-consistent field iterations

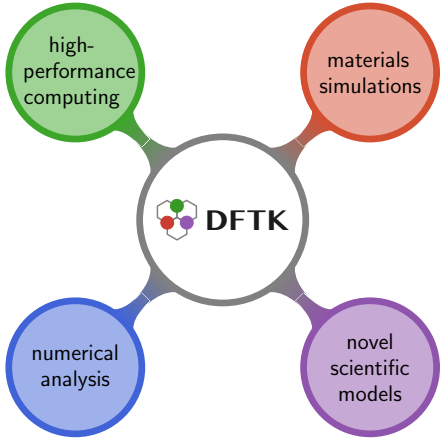
Michael F. Herbst*, Antoine Levitt

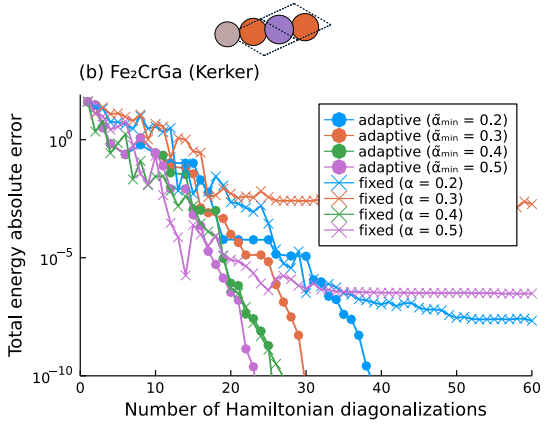
*Applied and Computational Mathematics, RWTH Aachen University
https://michael-herbst.com

15 September 2021

Slides: https://michael-herbst.com/talks/2021.09.15_youngcas_linesearch.pdf







Guaranteed & efficient line searches

a) Setting

- density matrix $\rho = \sum_{i=1}^{N_b} f_i |\phi_i\rangle\langle\phi_i|$
- Sets

$$\mathcal{H} = \{ H \in \mathbb{R}^{N_b \times N_b},\ H^T = H \}$$

$$\mathcal{P} = \{ \rho \in \mathcal{H},\ 0 < \rho < 1 \}$$

- Energy Functional $\mathcal{E}_0$

- KS Matrix : <u>Gradient</u> $H_{KS}(\rho) = \nabla \mathcal{E}_0(\rho)$

  related to $d\mathcal{E}_0(\rho) \cdot Q = \mathrm{tr}\left[H_{KS}(\rho)\, Q \right]$

- KS Kernel : <u>Hessian</u>

$$\underline{\underline{K}}(\rho) = \underline{\underline{d^2}}\mathcal{E}_0(\rho) = \underline{\underline{d\nabla}}\mathcal{E}_0(\rho)$$

- Fermionic entropy

$$s(\rho) = -\left[\rho \log \rho + (1-\rho) \log (1-\rho) \right]$$

$$s'(\rho) = \log\left(\frac{1-\rho}{\rho} \right) ; \quad s''(\rho) = -\frac{1}{\rho(1-\rho)}$$

- Free energy

$$\mathcal{E}(P) = \mathcal{E}_0(P) - \frac{1}{\beta} \mathrm{Tr}\left[s(P)\right] - \mu \,\mathrm{Tr}\, P$$

- $\nabla \mathcal{E}(P) = 0$ gives

$$0 = H_{KS}(P) - \mu - \frac{1}{\beta} s'(P)$$

$$\Rightarrow \boxed{P = f_{FD}(H_{KS}(P))} \; ; \; f_{FD}(H) = \frac{1}{1+e^{\beta(H-\mu)}}$$

- $\underline{\underline{\chi_0}} = d f_{FD}$ (Daleckii-Krein):

$$\underline{\underline{\chi_0}}\left(\sum_{i=1}^{N_b} \varepsilon_i |\phi_i\rangle\langle\phi_i|\right) \delta H = \sum_i^{N_b}\sum_j^{N_b} \frac{f_{FD}(\varepsilon_i) - f_{FD}(\varepsilon_j)}{\varepsilon_i - \varepsilon_j} \cdot |\phi_i\rangle\langle\phi_i, \delta H \phi_j\rangle\langle\phi_j|$$

- $$\underline{\underline{d^2\mathcal{E}(P)}} \cdot \delta P = -\frac{1}{\beta}\sum_i^{N_b}\sum_j^{N_b} \frac{s'(p_i) - s'(p_j)}{p_i - p_j} |\phi_j\rangle\langle\phi_j|\delta P \phi_i\rangle\langle\phi_i| + \underline{\underline{K}}(H_{KS}(P))$$

- $$\boxed{\underline{\underline{d^2\mathcal{E}(P)}} = \underline{\underline{K}}(H_{KS}(P)) + \underline{\underline{\Omega}}(f_{FD}^{-1}(P))}$$

where

$$\Omega\left[\sum_i^{N_b} \varepsilon_i |\phi_i\rangle\langle\phi_i|\right]\cdot\delta P = -\sum_{i}^{N_b}\sum_{j}^{N_b} \frac{\varepsilon_i - \varepsilon_j \quad |\phi_i\rangle\langle\phi_i|\delta P|\phi_j\rangle\langle\phi_j|}{f_{FD}(\varepsilon_i) - f_{FD}(\varepsilon_j)}$$

- Note: $\underline{\underline{\Omega}}(H) = -\left[\underline{\underline{\chi}}_0(H)\right]^{-1}$

---

b) "dual problem":

$$\mathcal{J}(H) = \mathcal{E}(f_{FD}(H))$$

$$\begin{aligned}
\nabla\mathcal{J}(H) &= \underline{\underline{\chi}}_0(H)\, \nabla\mathcal{E}(f_{FD}(H)) \\
&= \underline{\underline{\chi}}_0(H)\left[H_{KS}(f_{FD}(H)) - \mu - \frac{1}{\beta} S'\left[f_{FD}(H)\right]\right] \\
&= \underline{\underline{\chi}}_0(H)\left(H_{KS}(f_{FD}(H)) - H\right)
\end{aligned}$$

$$\underline{\underline{d^2\mathcal{J}}}(H_*) = -\underline{\underline{\chi}}_0(H_*)\left[1 - \underline{\underline{K}}(H_*)\right]\chi_0(H_*)$$

at stationary point $H_* = H_{KS}(f_{FD}(H_*))$

$\longrightarrow$

SCF versus direct minimisation:

- $H_{n+1} = H_{KS}(f_{FD}(H_n))$ "SCF"

$\Rightarrow$ gradient descent on $\mathcal{J}$ without damping preconditioned by $\underline{\underline{\chi}}_0(H_n) = -\underline{\underline{\Omega}}^{-1}(H_n)$

- $0 = \nabla \mathcal{E}(P_n) + \frac{1}{2} \underline{\underline{d^2\mathcal{E}}}(P_n) \cdot \delta P$

$$\Rightarrow \delta P = 2\left(\underline{\underline{\Omega}}(H_{KS}(P_n)) + \underline{\underline{K}}(f_{FD}^{-1}(P_n))\right)^{-1} \nabla \mathcal{E}(P_n)$$

Small gap $\Rightarrow$ $\underline{\underline{\Omega}}$ near-singular!

- $H_{n+1} = H_n + \alpha\left(H_{KS}(f_{FD}(H_n)) - H_n\right)$

converges for some α_* & $0 < \alpha < \alpha_*$

$\rightarrow$

d) Potential mixing (solid-state DFT)

- $\mathcal{E}_0(P) = \mathrm{Tr}(H_0 P) = g(\mathrm{diag}(P))$
- $H_{KS}(P) = H_0 + V(\mathrm{diag}(P))$

$$V(\rho) = \mathrm{diag}\,m(g'(\rho))$$

- $\rho(V) = \mathrm{diag}(f_{FD}(H_0 + V))$
- <u>Potential mixing:</u>

$$V_{n+1} = V_n + \alpha\left(V(\rho(V_n)) - V_n\right)$$

- preconditioned, damped iterations:

$$V_{n+1} = V_n + \alpha P^{-1}\left(V(\rho(V_n)) - V_n\right)$$

---

e) adaptive damping:

$$E_0(V_n + \alpha\,\delta V_n) = E_0(V_n) + \alpha\langle \nabla E_0(V_n), \delta V_n\rangle + \tfrac{1}{2}\alpha^2\langle \delta V_n, \nabla^2 E_0(V_n)\cdot\delta V_n\rangle + O(\alpha^3)$$

$$= E_0(V_n) + \alpha\langle \chi_0(V_n)\left[V(\rho(V_n)) - V_n\right], \delta V_n\rangle - \tfrac{1}{2}\alpha^2\langle \delta V_n, \chi_0(V_n)\left[1 - K(V_n)\chi_0(V_n)\right]\delta V_n\rangle + O$$

$$R_n = V(\rho(V_n)) - V_n \qquad \text{"residual"}$$

$$\chi_0 \, \delta V_n \simeq \delta\rho_n = \rho(V_n + \delta V_n) - \rho(V_n)$$

quadratic model:

$$\varphi_n(\alpha) = E_0(V_n) + \alpha \langle R_n, \delta\rho_n \rangle - \frac{1}{2}\alpha^2 \langle \delta\rho_n, \delta V_n - K(V_n)\,\delta\rho_n \rangle$$

<u>Algorithm:</u> V_n, $V_{search} = P^{-1} R_n$, trial damping $\tilde{\alpha}$

$\alpha_n \leftarrow \tilde{\alpha}$

loop

$V_{n+1} = V_n + \alpha_n V_{search}$

if accept V_{n+1} break

$\delta V_n \leftarrow \alpha_n V_{search}$

Build heuristic model φ_n

if model good

$\alpha_n \leftarrow \alpha_n \cdot \underset{\alpha}{\operatorname{argmin}}\ \varphi_n(\alpha)$

else

$\alpha_n \leftarrow \frac{\alpha_n}{2}$

end

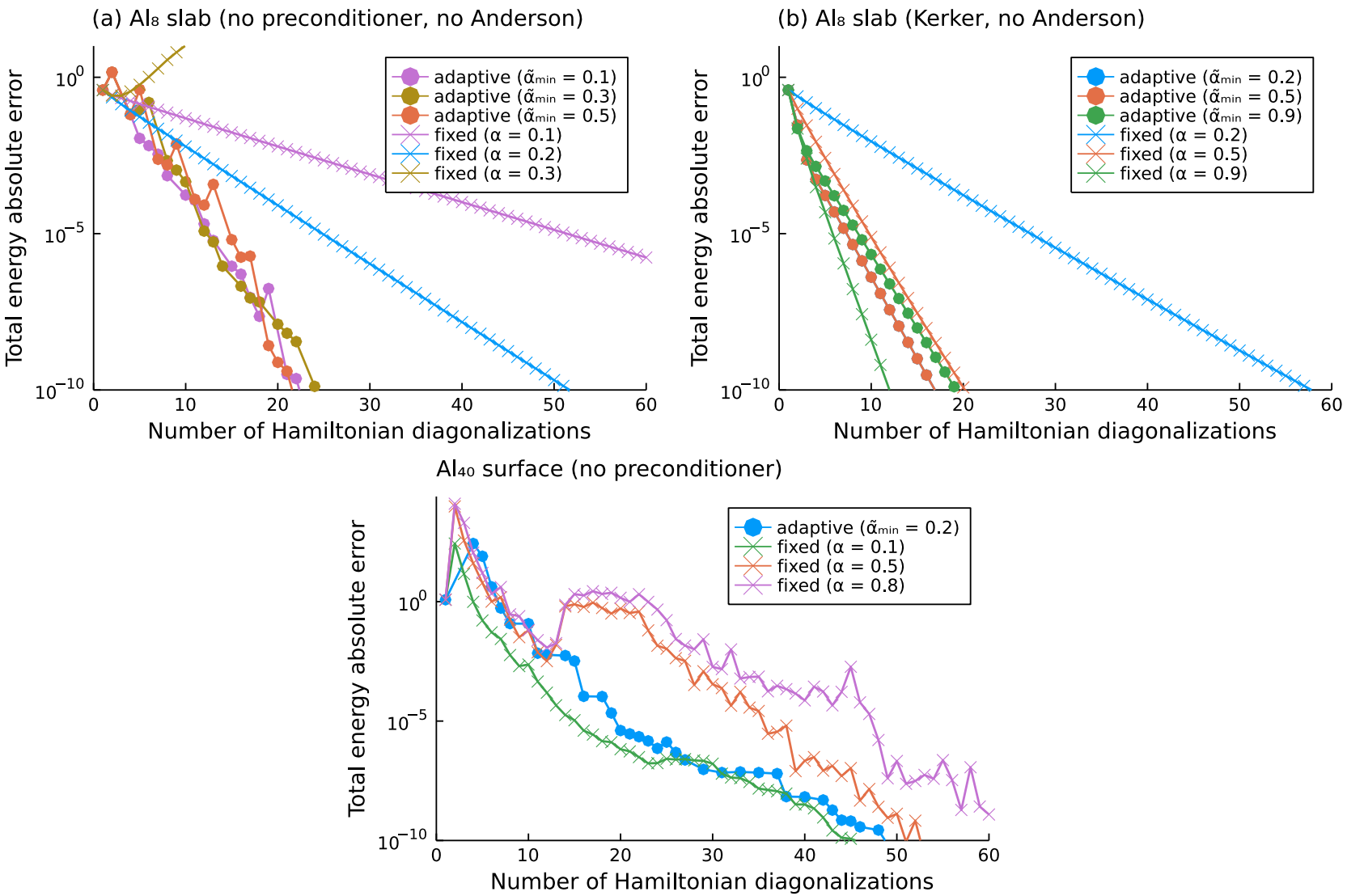
end

Results overview

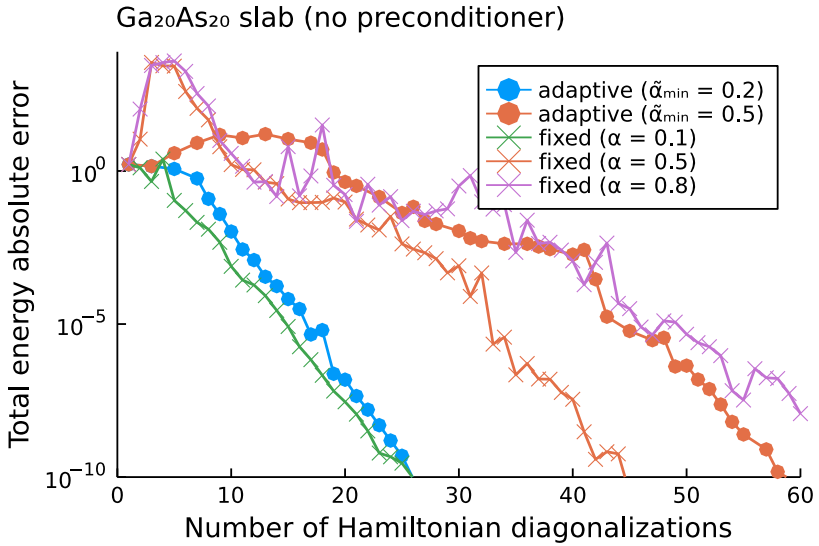
System	Precond.	fixed damping α										adaptive damping
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	
Al_8 slab	Kerker†	×	58	37	27	21	16	13	**11**	12	18	17
Al_8 slab	None†	×	**52**	×	×	×	×	×	×	×	×	24
Al_{40} slab	Kerker	19	15	14	12	**11**	12	12	12	12	12	12
Al_{40} slab	None	**38**	40	40	39	44	50	49	×	76	×	44
Al_{40} surface	None	**46**	48	50	49	51	60	61	66	89	×	49
$Ga_{20}As_{20}$ slab	None	**26**	33	40	42	45	44	70	70	65	76	26
CoFeMnGa	Kerker	×	×	×	×	28	**21**	24	28	22	22	30
Fe_2CrGa	Kerker	×	×	×	27	×	×	**19**	25	×	22	39
Fe_2MnAl	Kerker	×	48	×	×	×	20	21	17	16	**15**	34
$FeNiF_6$	Kerker	×	×	×	×	×	×	×	23	22	**21**	24
Mn_2RuGa	Kerker	×	×	×	×	37	24	23	**22**	23	23	36
Mn_3Si	Kerker	×	×	×	×	26	30	22	**20**	×	×	×

- PBE functional, realistic setup (Ecut, kgrid)
- Implementation and calculation in **DFTK** (`https://dftk.org`)

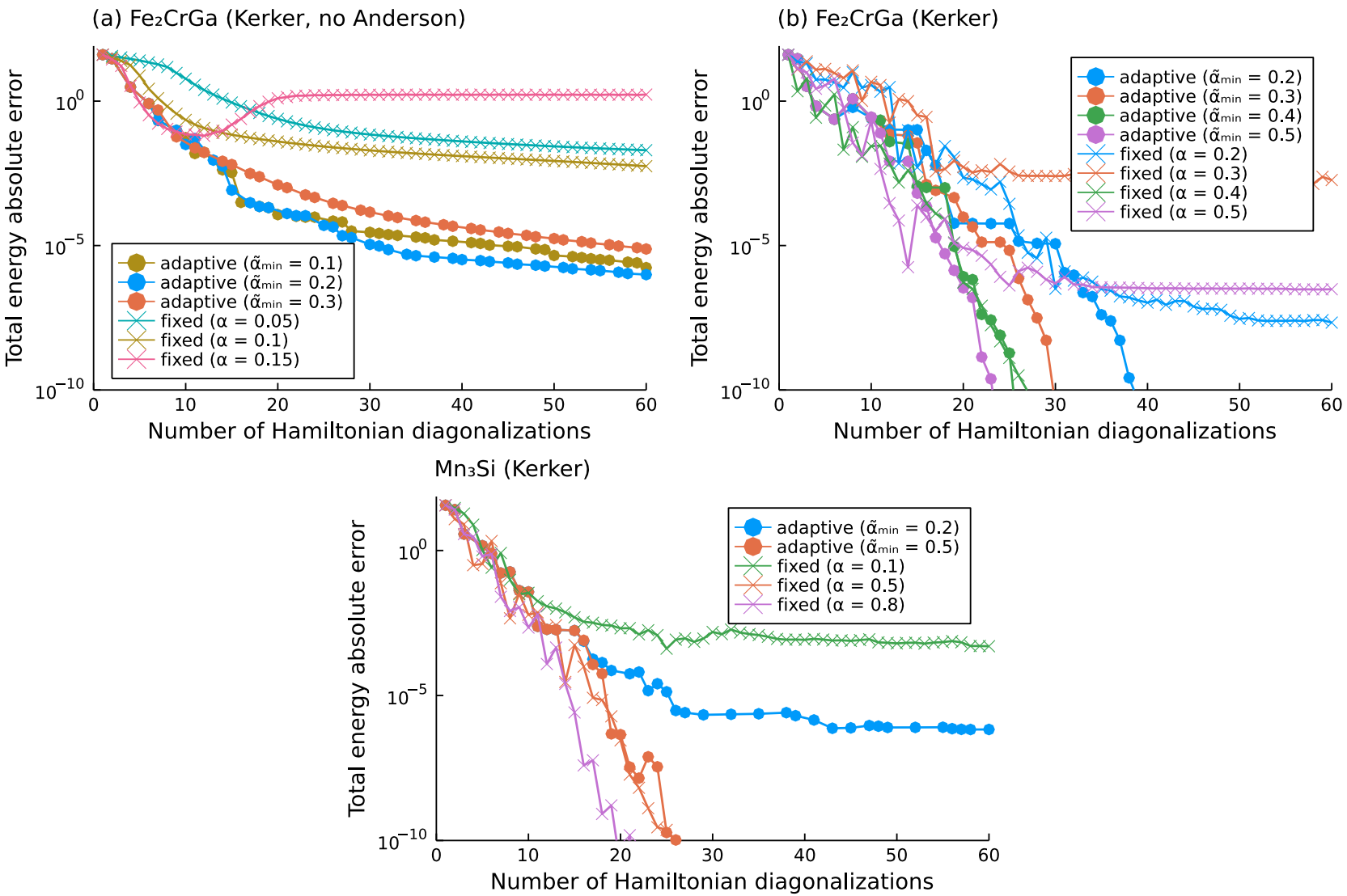
Aluminium slab and surface (simple systems)



GaAs slab (non-linear system)

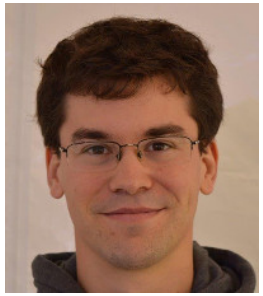


Fe_2CrGa and Mn_3Si (Heusler systems)



Acknowledgements

https://michael-herbst.com/talks/2021.09.15_youngcas_linesearch.pdf



Antoine Levitt

Benjamin Stamm

Eric Cancès

all DFTK contributors

Questions?

https://michael-herbst.com/talks/2021.09.15_youngcas_linesearch.pdf

DFTK https://dftk.org

julia: https://michael-herbst.com/learn-julia

mfherbst

https://michael-herbst.com/blog

herbst@acom.rwth-aachen.de

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International Licence.