Efficient sampling for lattice QMC

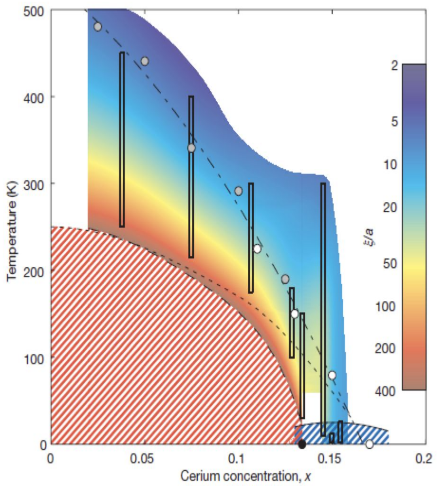
Joint work with Peter Lunts (UMD) and Michael Albergo (NYU)

Peter Lunts, Michael Albergo, and Michael Lindsey. **Non-Hertz-Millis scaling of the antiferromagnetic quantum critical metal via scalable Hybrid Monte Carlo**, *Nature Communications* 14 (2023).

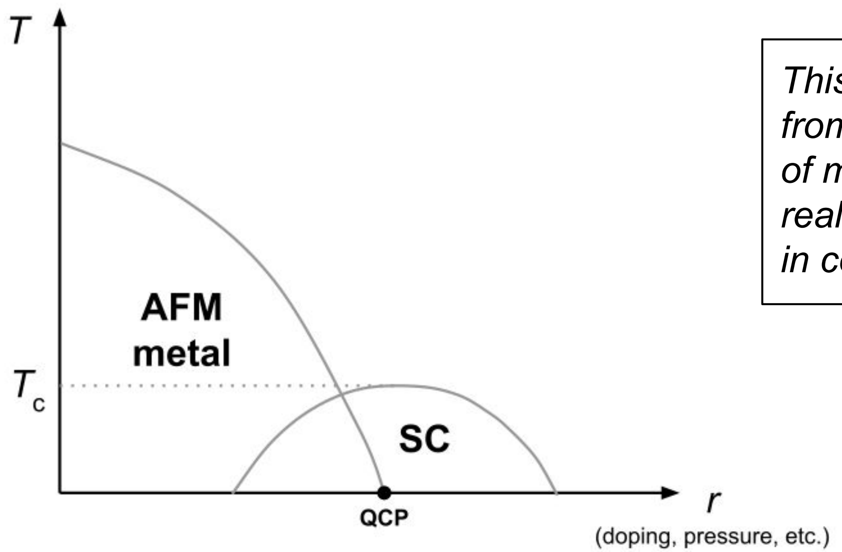
Motivation from high-T_c superconductivity

- Often arises in vicinity of antiferromagnetically (AFM) ordered phase of a metal
- *Ab initio* computations for known high-temperature superconducting materials are extraordinarily difficult
- There is broad interest in minimal effective models

e-doped cuprates



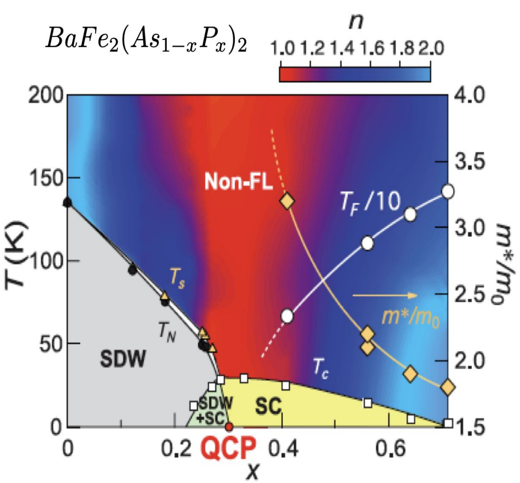
Motoyama et al., Nature (2007)



Cartoon phase diagram of a quantum critical metal

This picture emerges from experimental study of many quantum-critical real materials of interest in condensed matter

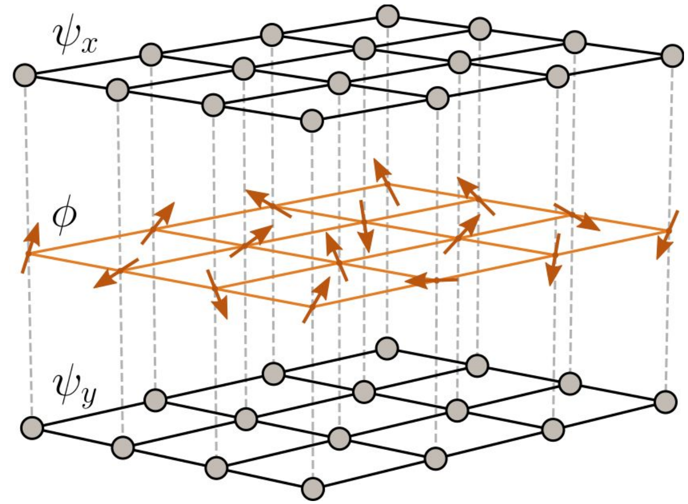
Fe-based superconductors



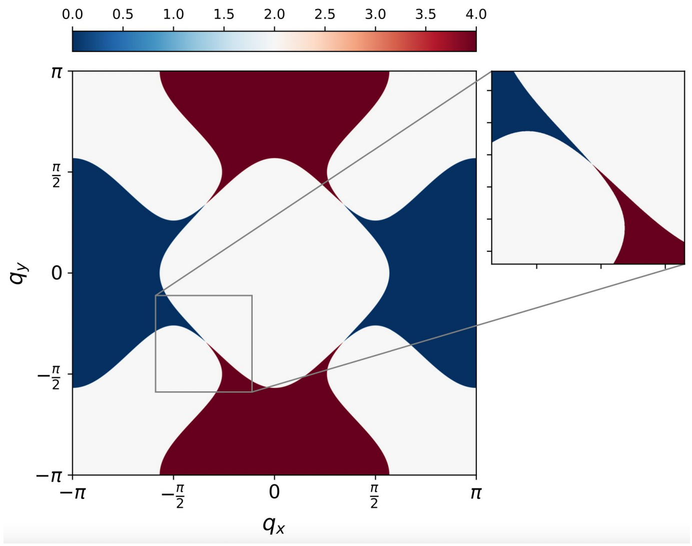
Hashimoto et al., Science (2012)

Spin-fermion model

- The O(3) spin-fermion model describes interactions between electrons in a metal and collective AFM spin excitations
  - Conjectured as model for high-T_c superconductivity near AFM metallic phase transition
- Key object is the 'bosonic' spin field ϕ
  - $\phi(x,y,\tau) \in R^3$, function of spatial lattice variables x, y and imaginary time variable τ
- Bosonic field couples to electrons
- Problem size parameters
  - L: lattice length
  - N: number of imaginary-time discretization points
- Total storage cost of ϕ is $O(L^2N)$



Two bands of fermions interacting with spin field, reproduced from [Bauer et al *Phys. Rev. Res.* (2020)]



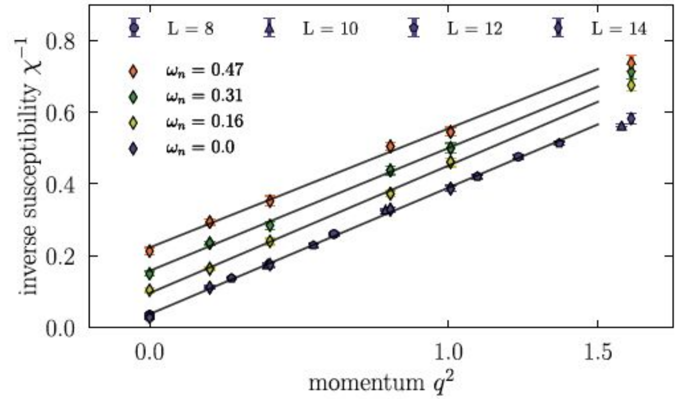
Fermi surface for the underlying noninteracting fermion model. (Two bands, two spins.)

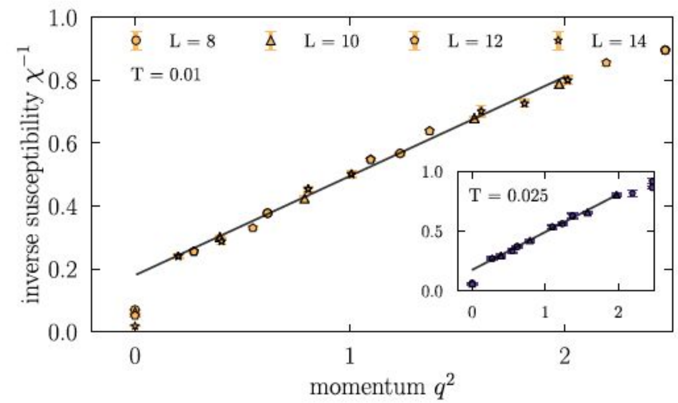
A computational challenge

- The key computational challenge is sampling bosonic fields according to the probability measure with density

$$P(\phi) \propto e^{-S_{\mathrm{B}}(\phi)} \det\left(M_{\phi}\right)$$

- Here M_{ϕ} is a positive definite operator of size $\sim L^2N$ admitting fast matrix-vector multiplications
  - Defines coupling of spin field to electrons
- Naive computational scaling of determinant is $O(L^6N^3)$
- Better scaling of $O(L^6N)$ is achieved by determinantal quantum Monte Carlo (DQMC)
  - Defines only existing competitive approach to spin-fermion model [Bauer et al *Phys. Rev. Res.* (2020)]





Inverse susceptibility of the spin field for two theory parameters. At bottom, **Hertz-Millis scaling** appears to fail, but correct scaling *cannot be resolved*!

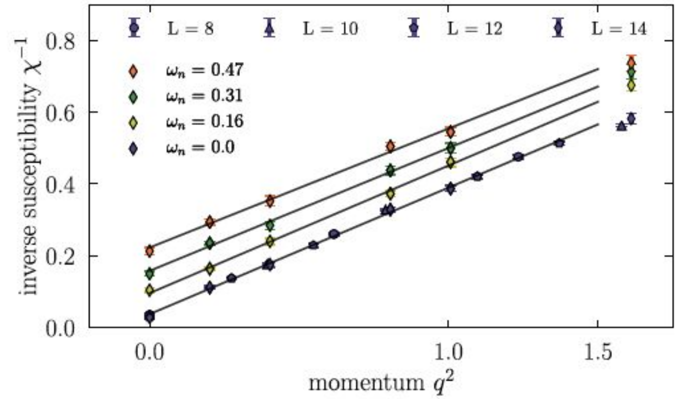
[Bauer et al *Phys. Rev. Res.* (2020)]

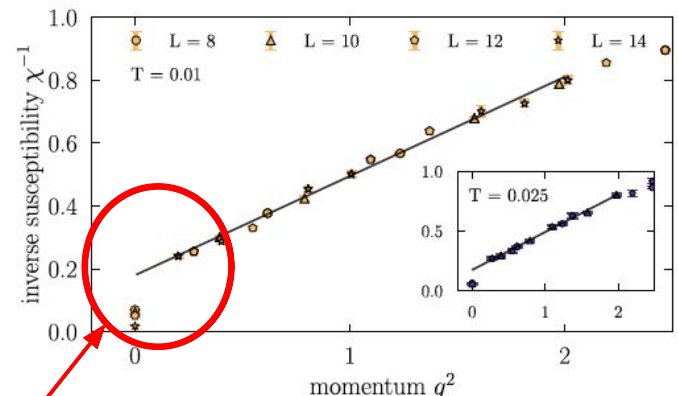
A computational challenge

- Finite-size model can only resolve physical correlation functions at small momenta of order $\boldsymbol{L}^{-1}$
  - Need larger lattice to reach experimentally accessible momenta
- DQMC can only reach $L \approx 14$, **not large enough to resolve novel physical scaling**
- In particular, want to test theoretical predictions of **non-Hertz-Millis scaling** of the spin correlation function

$$\chi(\omega, \boldsymbol{q}) = \int_0^\beta \sum_{\boldsymbol{r}} \langle \boldsymbol{\phi}(\tau, \boldsymbol{r}) \cdot \boldsymbol{\phi}(0, \boldsymbol{0}) \rangle e^{i\omega\tau - i\boldsymbol{q}\cdot\boldsymbol{r}} \, d\tau$$

  - Predicted in [Schlief, Lunts, and Lee, *Phys. Rev. X* (2017)] but theory is perturbative in a model parameter
  - Not reachable by DQMC





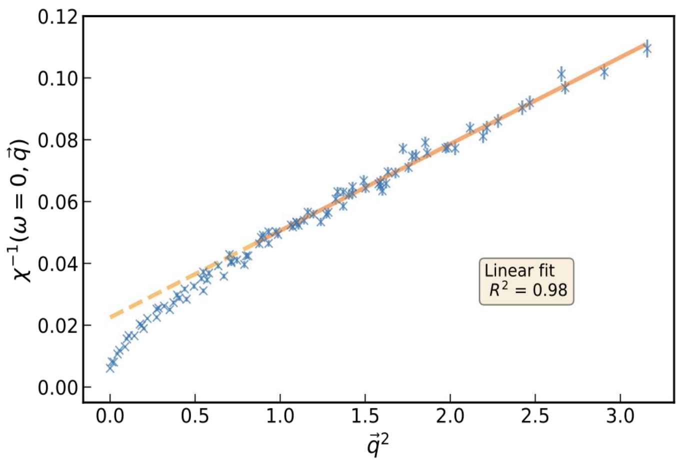
Inverse susceptibility of the spin field for two theory parameters. At bottom, **Hertz-Millis scaling** appears to fail, but correct scaling *cannot be resolved*!

[Bauer et al *Phys. Rev. Res.* (2020)]

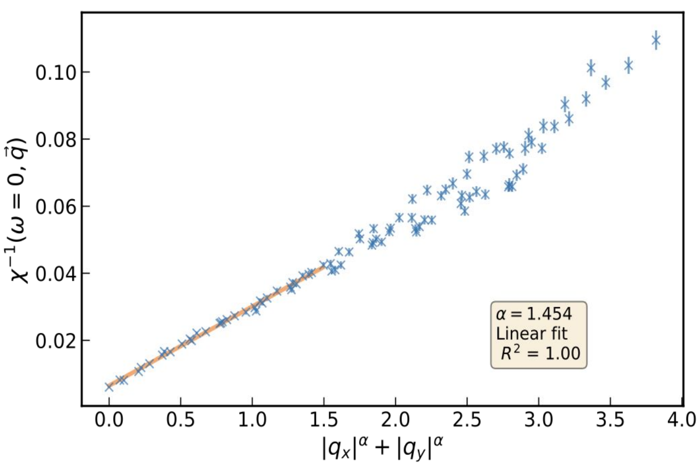
Missing data!

A computational challenge

- In the paper we go up to **80 x 80 x 200**. Some results:

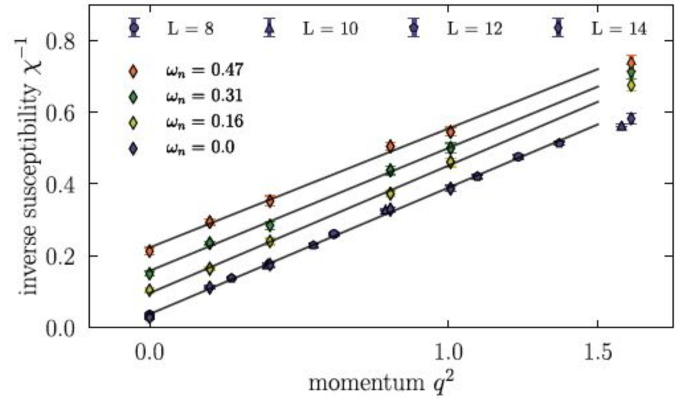


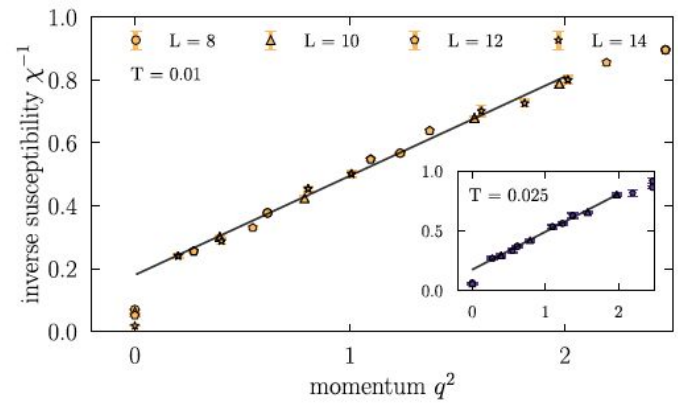
Confirmed failure of Hertz-Millis scaling



Determination of novel scaling

$$\chi^{-1}(\omega = 0, \vec{q}) \sim |q_x|^\alpha + |q_y|^\alpha$$





Inverse susceptibility of the spin field for two theory parameters. At bottom, **Hertz-Millis scaling** appears to fail, but correct scaling *cannot be resolved*!

[Bauer et al *Phys. Rev. Res.* (2020)]

Pseudofermion trick

- Borrowing a trick from the lattice quantum chromodynamics (LQCD) community, exchange the difficulty of computing determinants (rigid) for difficulty of solving linear systems (more flexible)
  - Trick is now widely used, e.g., by the group of Scalettar in other condensed matter settings
- Specifically, view

$$P(\phi) \propto e^{-S_{\mathrm{B}}(\phi)} \det\left(M_\phi\right)$$

  as the marginal of

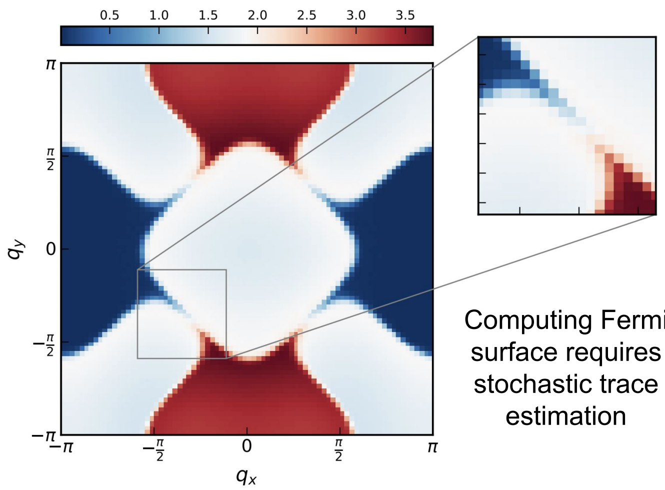
$$\widetilde{P}(\phi, \varphi) \propto e^{-S_{\mathrm{B}}(\phi) - \varphi^* M_\phi^{-1} \varphi}$$

- Sampling via Markov chain Monte Carlo (MCMC), specifically Hamiltonian Monte Carlo (HMC)

Discussion of the sampling problem

$$\widetilde{P}(\phi,\varphi) \propto e^{-S_{\mathrm{B}}(\phi)-\varphi^* M_\phi^{-1}\varphi}$$

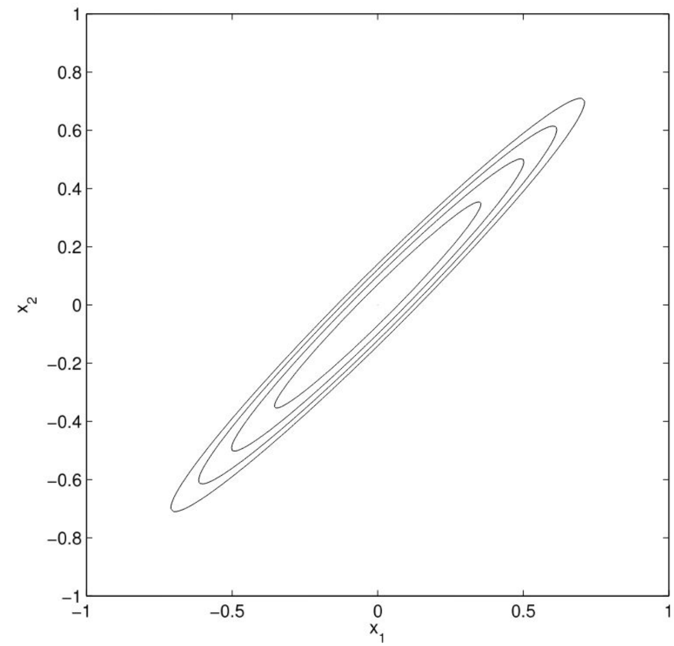
- Pseudofermion field φ can be updated efficiently with a Gibbs sampler step
  - Hence difficulty reduces to sampling from $\widetilde{P}(\,\cdot\,,\varphi)$
- Bottleneck per step of Markov chain is **solving linear systems** $M_\phi \chi = \varphi$ needed to evaluate the effective bosonic action $-\log \widetilde{P}(\,\cdot\,,\varphi)$ and its gradient
- Several obstacles to optimal $O(L^2N)$ scaling:
  - Preconditioning linear solves
    - Diagonal Fourier preconditioner works pretty well but need multigrid for true linear scaling in large *N* limit
  - Computation of fermionic observables
    - Requires the computation of diagonal of an inverse via stochastic trace estimation, higher order observables challenging to achieve with optimal scaling



Computing Fermi surface requires stochastic trace estimation

Correcting anisotropy

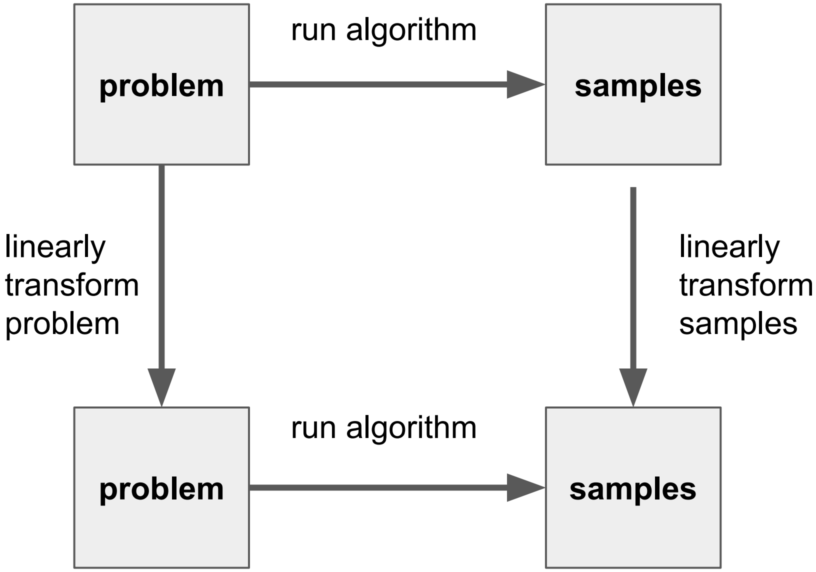
- But there is another hidden cost that threatens linear scaling: poor conditioning of the **distribution** itself!
    - Slows down the **mixing time** of Markov chain as the sampling problem becomes **high-dimensional** (i.e., *L* and *N* become large)
- Why HMC?
    - Among Metropolis-corrected schemes, best scaling of autocorrelation time ~ $d^{1/4}$ w.r.t dimension *d*
    - However, this scaling is derived for ***isotropic*** distributions
    - Need to correct anisotropic distributions
- Implicit in the specification of **<u>all</u>** local MCMC samplers (RWMH, Langevin, HMC) is some **metric** which defines the distribution of the noise process
    - Correcting the metric means correcting anisotropy



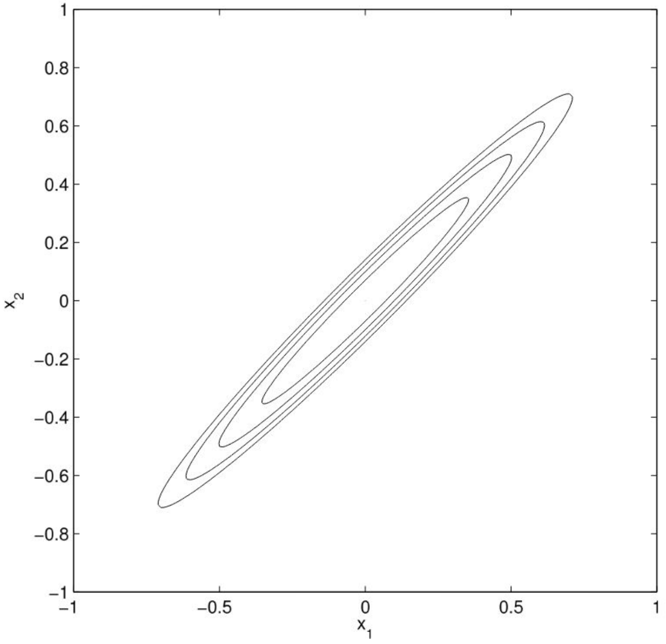
Example of an anisotropic probability density

Correcting anisotropy

- Inspired by affine-invariant samplers [Goodman and Weare], want the following commutative diagram to hold



- Unfortunately, generic affine-invariant samplers either suffer from curse of dimensionality or have $O(d^3)$ scaling (due to estimating and operating with the **covariance matrix**)



Example of an anisotropic probability density

Correcting anisotropy

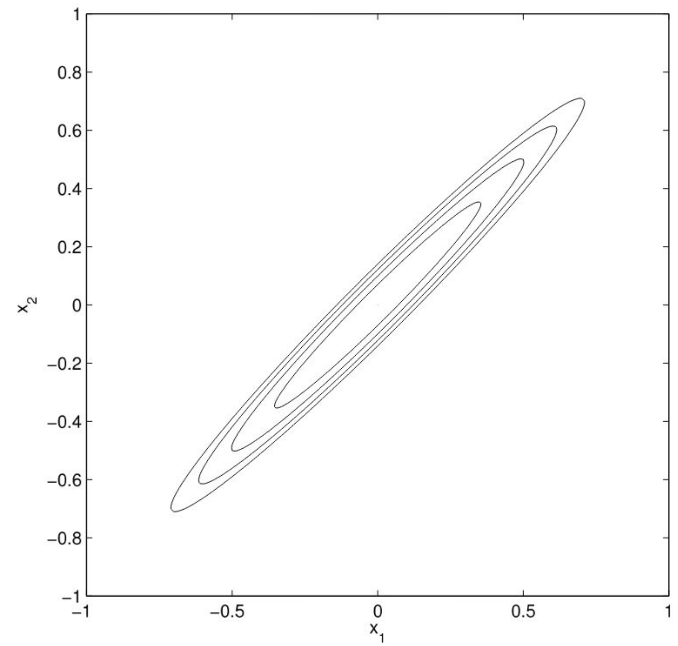
- However…the true covariance matrix Σ of the bosonic field is translation-invariant, hence admits fast linear operations via FFT
- Unfortunately Σ is unknown, and the non-interacting covariance is not good enough
- We must estimate it online with non-TI bosonic field samples $\phi^{(s)}$
- Empirical covariance (impractical to form)

$$\frac{1}{S} \sum_s \phi^{(s)} \phi^{(s)\top}$$

- We know that ground truth is diagonal in Fourier space, so we directly estimate the diagonal

$$\hat{\sigma} := \frac{1}{S} \sum_s \widehat{\phi}^{(s)} \odot \overline{\widehat{\phi}^{(s)}}$$

- Only $O(1)$ samples to estimate the entire diagonal



Example of an anisotropic probability density

- Form metric as $M = \Sigma^{-1}$
- Fast matvecs and fast samples from $\mathcal{N}(0, M)$

Other auto-tuning techniques

- Each HMC sample is produced by integrating Hamilton's equations for n_{leap} time steps of step size ε
- How to tune these parameters?
- Adopt best practices from statistics community
- Choose ε as large as possible subject to

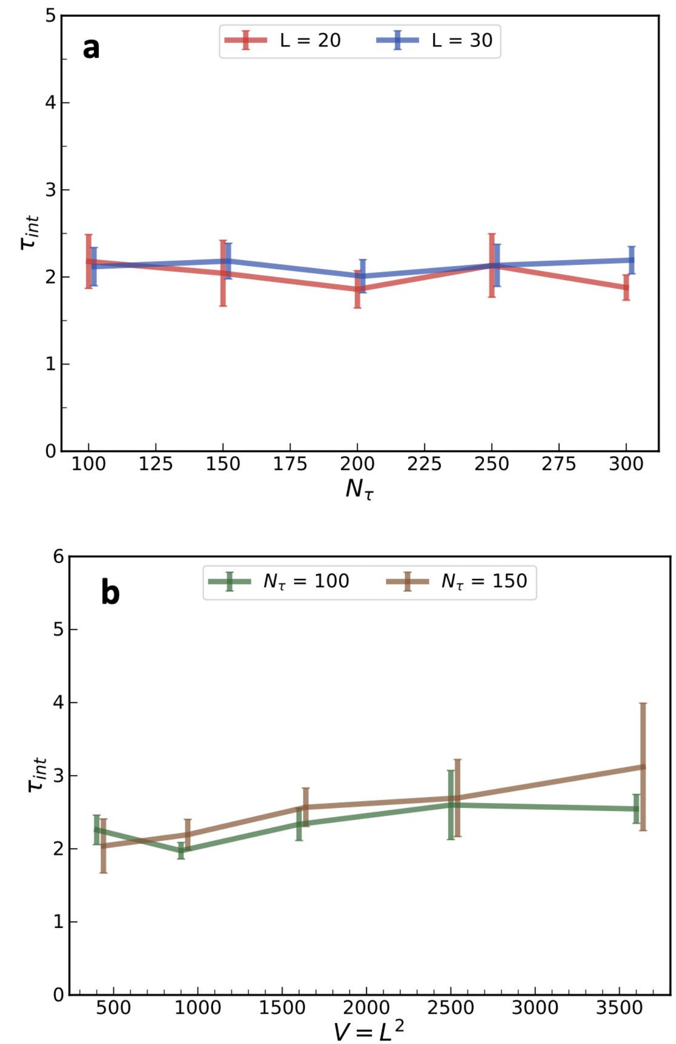
$$(1-\alpha(\varepsilon/2))^2 \leq 2(1-\alpha(\varepsilon))$$

  where $\alpha(\varepsilon)$ is expected acceptance rate of one step

- Choose n_{leap} to maximize

$$\mathbf{ESJD}(n) := \left\langle (\phi-\phi')^\top M(\phi-\phi')\,\alpha(\phi,\phi') \right\rangle_{\varepsilon, n_{\mathrm{leap}}=n}$$

[Pasarica and Gelman 2010]



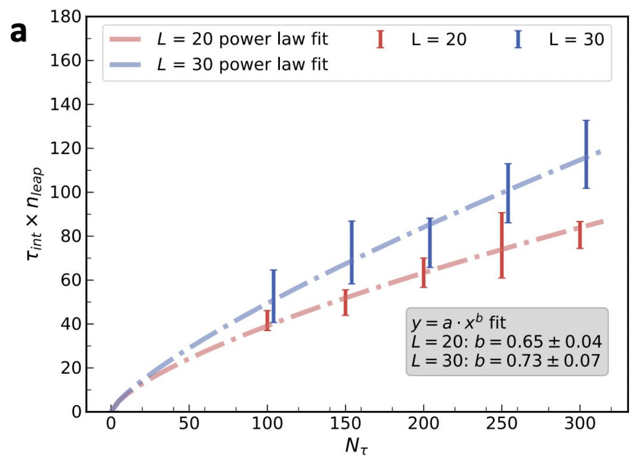
IAT as a function of (a) *N* and (b) *L*

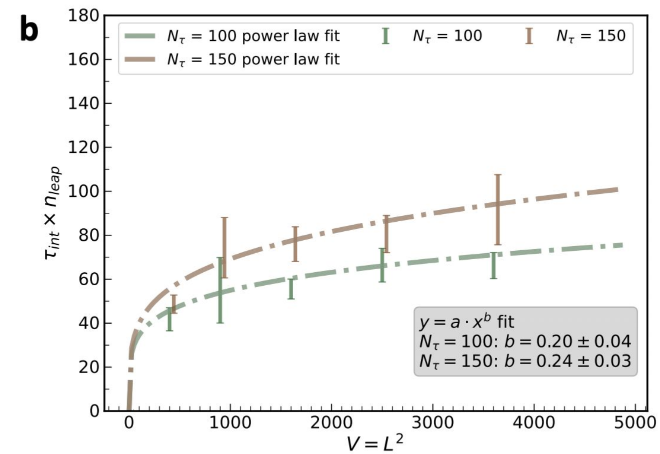
Scaling results (at critical point)

“Effective” autocorrelation time

Optimal HMC scaling of *b*=¼ with respect to volume.

Nearly optimal with respect to temperature.

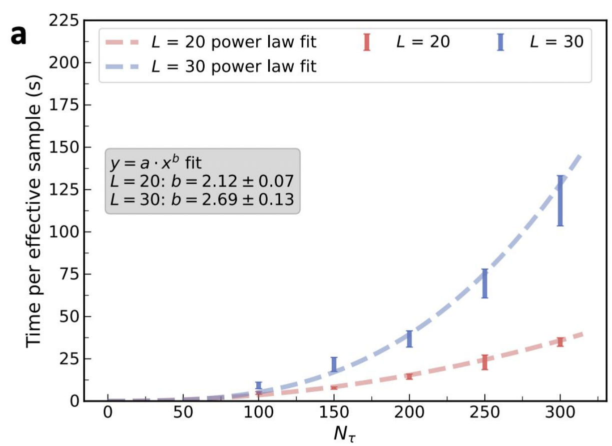


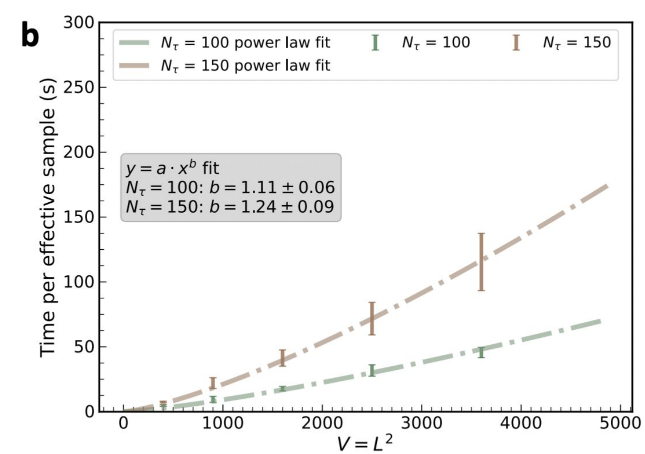


Wall clock scaling

Optimal with respect to volume

Suboptimal linear solver scaling with respect to temperature





Summary

- It's a good idea to tune your sampler!
  - Otherwise you are unlikely to get optimal scaling in high dimension.
- The tricks are quite portable.
- Some difficulties/questions:
  - How to construct efficient metric for gauge fields?
  - How to generalize notion of metric to complex Langevin?