The problem

We sample N-dimensional unit multivariate normal $q \sim \mathcal{N}(0, I_N)$ using mass matrix M and stepsize ϵ. These result in some single-step average acceptance probability $\langle a \rangle$.

Next we update the mass matrix to M_{new} and the goal is to find a new stepsize ϵ_{new} such that the new single-step acceptance probability is equal to the old one $\langle a_{\text{new}} \rangle = \langle a \rangle$.

We will only study the limit $\epsilon \to 0$, $a \to 1$. It may seem strange to take the limit $a \to 1$ here since we know that optimal tuning has $a \approx 0.8$. However, that result is for a calculated between the ends of a long trajectory but here a is calculated for a single-step only and is necessarily higher. We expect $a \to 1$ when $\epsilon \to 0$ is relevant (ie. for large N).

The integral

In the small stepsize limit the energy error of a single step starting from q with momemtum p is approximately

$$\Delta E(q, p) \approx \frac{\epsilon^3}{4} p^T M^{-2} q$$

The acceptance probability is related to the energy error by (using notation $\|x\|_- = \min(0, x)$)

$$\begin{aligned}
a(q, p) &= \min(1, \exp(\Delta E(q, p))) \\
&= \exp\left(\|\Delta E(q, p)\|_-\right) \\
&\approx 1 + \|\Delta E(q, p)\|_- \\
&\approx 1 + \frac{\epsilon^3}{4} \left\|p^T M^{-2} q\right\|_-
\end{aligned}$$

Let $\langle \cdot \rangle$ denote the expectation over $q \sim \mathcal{N}(0, I_N)$, $p \sim \mathcal{N}(0, M)$. Reparametrize $p = Lz$ with the cholesky factor $LL^T = M$ because then $z \sim \mathcal{N}(0, I_N)$. The average acceptance probability is

$$\begin{aligned}
\langle a(q, p) \rangle &\approx 1 + \frac{\epsilon^3}{4} \left\langle \left\|p^T M^{-2} q\right\|_- \right\rangle \\
&= 1 + \frac{\epsilon^3}{4} \left\langle \left\|z^T L^T M^{-2} q\right\|_- \right\rangle \\
&= 1 + \frac{\epsilon^3}{4} \left\langle \left\|z^T L^{-1} M^{-1} q\right\|_- \right\rangle \\
&= 1 - \frac{\epsilon^3}{8} \left\langle \left|z^T L^{-1} M^{-1} q\right| \right\rangle
\end{aligned}$$

The last line switches to ordinary absolute value because the distribution is symmetric.

The integral $\mathcal{E} = \left\langle \left| z^T L^{-1} M^{-1} q \right| \right\rangle$ is too complicated to evaluate exactly but we can still find useful upper and lower bounds.

An upper bound

An upper bound is straightforward. Let $X = L^{-1} M^{-1}$.

$$\begin{aligned}
\mathcal{E} &= \left\langle \left| z^T X q \right| \right\rangle \\
&= \left\langle \sqrt{\left(z^T X q\right)^2} \right\rangle \\
&< \sqrt{\left\langle \left(z^T X q\right)^2 \right\rangle} \\
&= \sqrt{\left\langle \left(z^T X q\right) \left(q^T X^T z\right) \right\rangle} \\
&= \sqrt{\left\langle \operatorname{trace}\left(X q q^T X^T z z^T\right) \right\rangle} \\
&= \sqrt{\operatorname{trace}\left(X \left\langle q q^T X^T z z^T \right\rangle\right)}
\end{aligned}$$

Recall that q and z are independent normal variates.

$$\begin{aligned}
\mathcal{E} &< \sqrt{\operatorname{trace}\left(X \left\langle q q^T \right\rangle X^T \left\langle z z^T \right\rangle\right)} \\
&= \sqrt{\operatorname{trace}\left(X I_N X^T I_N\right)} \\
&= \sqrt{\operatorname{trace}\left(X X^T\right)} \\
&= \sqrt{\operatorname{trace}\left(M^{-3}\right)} \\
&= \left\| M^{-\frac{3}{2}} \right\|_F
\end{aligned}$$

Here $\|\cdot\|_F$ is the Frobenius norm.

Lower bounds

Let's start by taking the singular value decomposition $X = U \Sigma V$. The rotation matrices U and V are cancelled by the rotational invariance of z and q.

$$\mathcal{E} = \left\langle \left| z^T U \Sigma V q \right| \right\rangle = \left\langle \left| z^T \Sigma q \right| \right\rangle = \left\langle \left| \sum_i \sigma_i z_i q_i \right| \right\rangle$$

Consider $\mathcal{E}$ as a function of the singular values (in descending order $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_N$). This function is homegenous, convex and monotonically increasing.

$$\begin{aligned}
\mathcal{E} &= \mathcal{E}\left(\sigma_1, \sigma_2, \cdots, \sigma_N\right) \\
&= \sigma_K \mathcal{E}\left(\frac{\sigma_1}{\sigma_K}, \cdots, \frac{\sigma_K}{\sigma_K}, \frac{\sigma_{K+1}}{\sigma_K}, \cdots, \frac{\sigma_N}{\sigma_K}\right) \\
&\geq \sigma_K \mathcal{E}\left(\underbrace{1, \cdots, 1}_{K}, \underbrace{0, \cdots, 0}_{N-K}\right) \\
&= \sigma_K \left\langle \left| \sum_{i=1}^{K} z_i q_i \right| \right\rangle \\
&= \sigma_K \left\langle \left| z \cdot q \right| \right\rangle = \sigma_K \left\langle |z| \left| q_z \right| \right\rangle
\end{aligned}$$

The vectors z and q are here taken to be K-dimenstional. The projection q_z has 1D unit normal distribution and is independent of $|z|$, which has χ distribution with K degrees of freedom. All we need to know that $\langle |z| \rangle \langle |q_z| \rangle > \frac{1}{2}\sqrt{K}$. This gives us lower bounds

$$\forall k : \mathcal{E} > \frac{1}{2}\sqrt{k}\sigma_k$$

Recall the upper bound $\left\| M^{-\frac{3}{2}} \right\|_F = \sqrt{\sum_k \sigma_k^2}$. How far can it be from the highest lower bound?

The gap between the upper and the highest lower bound is at it's maximum when all lower bounds are equal. This must be so because if there was a k such that the kth lower bound $\frac{1}{2}\sqrt{k}\sigma_k$ is below the highest lower bound then you could increase the upper bound without changing the highest lower bound by increasing the value of σ_k.

Thus, the maximum gap happens when $\sigma_k = \sqrt{k}^{-1}\sigma_1$ and it is

$$\frac{\sqrt{\sum_k \sigma_k^2}}{\max_k \left(\frac{1}{2}\sqrt{k}\sigma_k\right)} = \frac{\sqrt{\sum_k \left(\sqrt{k}^{-1}\sigma_1\right)^2}}{\frac{1}{2}\sigma_1} = 2\sqrt{\sum_{k=1}^{N} k^{-1}} \leq 2\sqrt{1 + \log N}$$

We have a general lower bound

$$\mathcal{E} > \frac{1}{2\sqrt{1 + \log N}} \left\| M^{-\frac{3}{2}} \right\|_F$$

Putting it all together

Let's get back to $\langle a \rangle = \langle a_{\text{new}} \rangle$.

$$\begin{aligned}
\langle a\rangle &= \langle a_{\text{new}}\rangle \\
1-\frac{\epsilon^3}{8}\mathcal{E} &= 1-\frac{\epsilon_{\text{new}}^3}{8}\mathcal{E}_{\text{new}} \\
\epsilon^3\mathcal{E} &= \epsilon_{\text{new}}^3\mathcal{E}_{\text{new}}
\end{aligned}$$

Now apply the bounds

$$\begin{aligned}
\epsilon^3\frac{1}{2\sqrt{1+\log N}}\left\|M^{-\frac{3}{2}}\right\|_F &< \epsilon_{\text{new}}^3\left\|M_{\text{new}}^{-\frac{3}{2}}\right\|_F \\
\epsilon^3\left\|M^{-\frac{3}{2}}\right\|_F &> \epsilon_{\text{new}}^3\frac{1}{2\sqrt{1+\log N}}\left\|M_{\text{new}}^{-\frac{3}{2}}\right\|_F
\end{aligned}$$

These can be rearranged to

$$\sqrt[3]{\frac{1}{2\sqrt{1+\log N}}\left\|M^{-\frac{3}{2}}\right\|_F\left\|M_{\text{new}}^{-\frac{3}{2}}\right\|_F^{-1}} < \frac{\epsilon_{\text{new}}}{\epsilon} < \sqrt[3]{2\sqrt{1+\log N}\left\|M^{-\frac{3}{2}}\right\|_F\left\|M_{\text{new}}^{-\frac{3}{2}}\right\|_F^{-1}}$$

NB: $\left\|M^{-\frac{3}{2}}\right\|_F\left\|M_{\text{new}}^{-\frac{3}{2}}\right\|_F^{-1} \leq \left\|M^{-\frac{3}{2}}M_{\text{new}}^{\frac{3}{2}}\right\|_F$ because the norm is submultiplicative.