Interactive Sound Propagation with Bidirectional Path Tracing Supplementary Material

1 Convergence of SNR optimization

In our propagation algorithm, we need to optimize the SNR metric

$$\sum_{m=1}^{M} 5\log_{10}(\sum_{n=1}^{N} \frac{\sigma_{mn}^2}{Sx_n}). \tag{1}$$

where x_n is the sample probability for the integral $T^i L_0$. Optimization of this target function above could be written as

$$\min_{\mathbf{x}\in\mathbb{R}_+^N} f(\mathbf{x}) \text{ s.t. } \sum_{n=1}^{N} x_n = 1 \tag{2}$$

where

$$f(\mathbf{x}) = \sum_{m=1}^{M} \ln(\sum_{n=1}^{N} \frac{a_{mn}}{x_n}), a_{mn} \geq 0 \tag{3}$$

with at least one positive a_{mn} for every m and n. This optimization problem is strictly convex, which guarantees the existence and uniqueness of the global minimum. To find the solution for this problem, one could use the iterative method below:

$$\mathbf{x}_{i+1} = \alpha\mathbf{T}(\mathbf{x}_i) + (1-\alpha)\mathbf{x}_i, \tag{4}$$

where

$$T_n(\mathbf{x}) = \frac{1}{M}\sum_{m=1}^{M} \frac{\frac{a_{mn}}{x_n}}{\sum_{k=1}^{N}\frac{a_{mk}}{x_k}} \tag{5}$$

. In this section, we'll prove that the algorithm above locally converges to the global minimum for every $\alpha \in (0,1)$ and the convergence is at least linear.

The construction of our iterative method starts from the method of Lagrange multipliers [Bertsekas and Nedic 2003]. The correspondent Lagrange function for (2) is

$$\mathcal{L}(\mathbf{x},\lambda) = f(\mathbf{x}) + \lambda(\sum_{n=1}^{N} x_n - 1) \tag{6}$$

and the solution $\mathbf{x}^*$ satisfies the equation $\nabla\mathcal{L}(\mathbf{x}^*,\lambda^*) = 0$ for some certain λ^*. The equation could also be written as

$$\begin{cases} \frac{\partial f}{\partial x_n}(\mathbf{x}^*) = -\lambda^* \\ ||\mathbf{x}^*|| = 1 \end{cases}. \tag{7}$$

Given

$$\frac{\partial f}{\partial x_n} = -\sum_{m=1}^{M} \frac{\frac{a_{mn}}{x_n^2}}{\sum_{k=1}^{N}\frac{a_{mk}}{x_k}}, \tag{8}$$

(8) is a set of nonlinear equations, which is hard to solve directly. Therefore, we constructed a iterative method to obtain the approximation of the solution $\mathbf{x}^*$. For an iteration method $\mathbf{x}_{n+1} = \mathbf{T}(\mathbf{x}_n)$ that solves (8), it's necessary that $\mathbf{T}(\mathbf{x})$ maps $\mathbf{x}^*$, the solution of (8), to itself. And $\mathbf{x}^*$ is called a "fixed point" of $\mathbf{T}(\mathbf{x})$.

We have constructed a $\mathbf{T}(\mathbf{x})$ which has a fixed point at $\mathbf{x}^*$:

$$\mathbf{T}(\mathbf{x}^*) = \frac{\lambda^*\mathbf{x}^*}{M}. \tag{9}$$

We could see from (5) that $||\mathbf{T}(\mathbf{x}^*)|| = 1$. Together with (7), we have $\lambda^* = M$. Thus $\mathbf{x}^*$ is a fixed point of operator $\mathbf{T}$. With

$$T_n(\mathbf{x}) = -\frac{x_n}{M}\frac{\partial f}{\partial x_n} \tag{10}$$

we have the equation (5). To further adjust the convergence of the iteration method, we add a relaxation factor α to $\mathbf{T}(\mathbf{x})$ and achieve the iteration algorithm (4).

Now we need to prove the convergence of our algorithm. First we'll look at the Jacobian matrix of operator $\mathbf{T}$ at $\mathbf{x}^*$. We have

$$\frac{\partial T_i}{\partial x_i} = \frac{1}{Mx_i}\sum_{m=1}^{M}\left(-\frac{\frac{a_{mi}}{x_i}}{\sum_{k=1}^{N}\frac{a_{mk}}{x_k}} + \frac{\frac{a_{mi}^2}{x_i^2}}{(\sum_{k=1}^{N}\frac{a_{mk}}{x_k})^2}\right) \tag{11}$$

and

$$\frac{\partial T_i}{\partial x_j} = \frac{1}{Mx_i}\sum_{m=1}^{M}\frac{\frac{a_{mi}a_{mj}}{x_j^2}}{(\sum_{k=1}^{N}\frac{a_{mk}}{x_k})^2}, i \neq j. \tag{12}$$

We know from (7) that

$$\frac{1}{Mx_i^*}\sum_{m=1}^{M}\frac{\frac{a_{mi}}{x_i^*}}{\sum_{k=1}^{N}\frac{a_{mk}}{x_k^*}} = \frac{\lambda^*}{M} = 1, \tag{13}$$

And the Jacobian matrix of $\mathbf{T}$ at $\mathbf{x}^*$ could be expressed as

$$D\mathbf{T}(\mathbf{x}^*) = \mathbf{A} - \mathbf{I}, \tag{14}$$

where $\mathbf{I}$ is the identity matrix and

$$\mathbf{A} = [a_{ij}]_{N\times N}, a_{ij}(\mathbf{x}) = \frac{1}{M}\sum_{m=1}^{M}\frac{\frac{a_{mi}a_{mj}}{x_i x_j^2}}{(\sum_{k=1}^{N}\frac{a_{mk}}{x_k})^2}. \tag{15}$$

Now we'll prove that all the eigenvalues of $\mathbf{A}(\mathbf{x}^*)$ are inside the $[0,1]$ interval. First, we notice that

$$\mathbf{A} = \mathbf{BC}, \tag{16}$$

where

$$\mathbf{B}(\mathbf{x}) = \left[\frac{\frac{a_{mi}}{x_i}}{\sum_{k=1}^{N}\frac{a_{mk}}{x_k}}\right]_{N\times M} \tag{17}$$

$$\mathbf{C}(\mathbf{x}) = \left[\frac{\frac{a_{mj}}{x_j^2}}{M\sum_{k=1}^{N}\frac{a_{mk}}{x_k}}\right]_{M\times N} \tag{18}$$

It is obvious that $||\mathbf{B}||_1 = 1$. Further, we know from (13) that $||\mathbf{C}(\mathbf{x}^*)||_1 = 1$. Therefore we get $||\mathbf{A}(\mathbf{x}^*)||_1 \leq 1$. Since the spectral radius of a matrix is no greater than its norm, the upper bound of the eigenvalue is proved.

Second, we could also write $\mathbf{A}$ as

$$\mathbf{A} = \frac{1}{M}\mathbf{B}^T\mathbf{B}\mathbf{\Lambda}^{-1} = \frac{1}{M}\sqrt{\mathbf{\Lambda}}(\sqrt{\mathbf{\Lambda}}^{-1}\mathbf{B}^T\mathbf{B}\sqrt{\mathbf{\Lambda}}^{-1})\sqrt{\mathbf{\Lambda}}^{-1} \tag{19}$$

where $\mathbf{\Lambda}(\mathbf{x}) = \text{diag}(\mathbf{x})$. We see that $M\cdot\mathbf{A}$ is similar to the matrix $\sqrt{\mathbf{\Lambda}}^{-1}\mathbf{B}^T\mathbf{B}\sqrt{\mathbf{\Lambda}}^{-1}$, which is a positive semidefinite matrix with no negative eigenvalue. Thus all eigenvalues of $\mathbf{A}$ are non-negative and the lower bound of the eigenvalue is proved.

Combining the conclusion above with (14), we know that all the eigenvalues of $D\mathbf{T}(\mathbf{x}^*)$ fall into the interval $[-1, 0]$. Now we'll do a first-order Taylor expansion of $\mathbf{T}$ at $\mathbf{x}^*$:

$$\mathbf{T}(\mathbf{x}) = \mathbf{x}^* + D\mathbf{T}(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) + \mathbf{E}(\mathbf{x})(\mathbf{x} - \mathbf{x}^*) \tag{20}$$

where $\mathbf{E}(\mathbf{x})$ is a matrix that satisfies

$$\lim_{\mathbf{x} \to \mathbf{x}^*} ||\mathbf{E}(\mathbf{x})|| = 0. \tag{21}$$

Together with (4), we have

$$\begin{aligned} \frac{||\mathbf{x}_{i+1} - \mathbf{x}^*||}{||\mathbf{x}_i - \mathbf{x}^*||} &\leq ||\alpha(D\mathbf{T}(\mathbf{x}^*) + \mathbf{E}(\mathbf{x})) + (1-\alpha)\mathbf{I}|| \\ &\leq \max\{|1 - 2\alpha|, |1 - \alpha|\} + \delta(||\mathbf{x} - \mathbf{x}^*||) \end{aligned} \tag{22}$$

with $\lim_{x \to 0^+} \delta(x) = 0$.

For any $\alpha \in (0, 1)$, we can always find an ε_0 that satisfies $\max\{|1 - 2\alpha|, |1 - \alpha|\} + \delta(x) < 1$ for any $x < \varepsilon_0$. Therefore (4) linearly converges to $\mathbf{x}^*$ for any initial $\mathbf{x}$ with $||\mathbf{x} - \mathbf{x}^*|| < \varepsilon_0$. Notice that $\max\{|1 - 2\alpha|, |1 - \alpha|\}$ reaches its minimum when $\alpha = \frac{2}{3}$, which gives us a good candidate (but not necessarily the best choice) for α in practical applications.

2 Improving Variance Estimation with Temporal Coherence

Our iteration algorithm requires an estimation of σ^2_{mn}. This estimation must be reevaluated constantly to address the changes of the sound environment. However, the variance estimation can be inaccurate due to insufficient number of samples. Actually, given a random variable X, we have

$$\sigma^2[S^2(X)] = \frac{1}{N}(\mu_4[X] - \frac{N-3}{N-1}\sigma^4[X]), \tag{23}$$

where S^2 is the sample variance estimation with N samples and μ_4 is the fourth central moment [Casella and Berger 2002].

We exploit the temporal coherence and combine the estimation of the current frame with the results from the previous frames to improve the estimation quality. From (23) we observe that $\sigma^2[S^2(X)]$ is roughly inversely proportional to N. To simplify our analysis, we would use a approximated version of (23), $\sigma^2[S^2(X)] = C/N$ in our following discussion.

We tag every estimated variance S^2 with a quality indicator $Q[S^2]$. which satisfies

$$\sigma^2[S^2] = \frac{C}{Q[S^2]} \tag{24}$$

One could see from the definition that an estimation with a larger Q will have less variance. For a new estimation, $Q[S^2]$ equals to the number of samples used for estimation. For the combination of two estimations, we have the equation below.

$$\begin{aligned} \sigma^2[\gamma S_a^2 + (1-\gamma)S_b^2] &= \gamma^2\sigma^2[S_a^2] + (1-\gamma)^2\sigma^2[S_b^2] \\ &= C\left(\frac{\gamma^2}{Q[S_a^2]} + \frac{(1-\gamma)^2}{Q[S_b^2]}\right) \end{aligned} \tag{25}$$

and therefore

$$Q[\gamma S_a^2 + (1-\gamma)S_b^2] = \frac{Q[S_a^2]Q[S_b^2]}{(1-\gamma)^2 Q[S_a^2] + \gamma^2 Q[S_b^2]} \tag{26}$$

It's not hard to see that $Q[\gamma S_a^2 + (1-\gamma)S_b^2]$ reaches its maximum at

$$\gamma = Q[S_{i-1}^2]/(Q[S_{i-1}^2] + Q[S_0^2]) \tag{27}$$

And its maximum value is $Q[S_a^2] + Q[S_b^2]$.

After we calculate a new estimation S_0^2 from samples in current frame, we will combine it with the estimation inherited from the last frame S_{i-1}^2 and generate the estimation of the current frame S_i^2. However, we also need to address the changes of the sound environment, which requires us to lower the weight of S_{i-1}^2 as much as possible. To balance between the quality and the responsiveness to scene changes, we use a predefined quality standard Q^*, and evaluate S_k^2 with the following rules:

- if $Q[S_0^2] > Q^*$, we'll use S_0^2 for S_k^2 directly;
- if $Q[S_0^2] + Q[S_{i-1}^2] < Q^*$, the maximal value of $Q[\gamma S_a^2 + (1-\gamma)S_b^2]$ would still be smaller than Q^*, and we'll use the optimal combination weight from (27);
- Otherwise, we would keep $Q[\gamma S_{i-1}^2 + (1-\gamma)S_0^2] = Q^*$ to make the quality of the estimation stable. The combination weight is achieved by solving this equation, whose solutions are given below:

$$\gamma = \frac{Q_{i-1} \pm \sqrt{Q_{i-1}Q_0(\frac{Q_{i-1}+Q_0}{Q^*} - 1)}}{Q_{i-1} + Q_0} \tag{28}$$

When $Q_{i-1} = Q^*$ (which is very likely to happen in practice), the equatiion above could be further simplified to:

$$\gamma = \frac{Q^* \pm Q_0}{Q^* + Q_0} \tag{29}$$

Since we need to lower the weight of S_{i-1}^2. we would choose the smaller value for γ.

References

BERTSEKAS, D., AND NEDIC, A. 2003. Convex analysis and optimization (conservative).

CASELLA, G., AND BERGER, R. L. 2002. *Statistical inference*, vol. 2. Duxbury Pacific Grove, CA.