A. Proofs

The RHS (10) is simply a linear objective of the form $\sum_i \alpha_i w_i$. The constraint set $w \in \mathbb{S}_+^{P-1}$ implies that $\|w\|_1 \geq 1$ and consequently, $\sum_i w_i \geq 1$. Therefore, the following inequalities hold

$$\sum_i \alpha_i w_i \geq (\min_i \alpha_i) \sum_i w_i \geq \min_i \alpha_i, \tag{26}$$

with equality when $a = e_{i^*}$, the one-hot vector corresponding to $i^* = \arg\min_i \alpha_i$. Any mixture is necessarily bigger by triangle inequality and because the problem is convex, e_{i^*} is a *unique* minimizer corresponding to the column of $W^{(K)}$ with the *smallest* norm.

The LHS (10) is identical to a quadratic objective, which we show has the same minimizer. Due to the non-negativity of W, dot products between columns of W must be non-negative. Subsequent inequalities follow

$$\begin{aligned} w^T G w &= \sum_i w_i^2 \|W_{:,i}\|_2^2 + 2 \sum_{i<j} w_i w_j W_{:,i}^T W_{:,j} \\ &\geq \sum_i w_i^2 \|W_{:,i}\|_2^2 \\ &\geq (\min_i \|W_{:,i}\|_2^2) \sum_i w_i^2 = \min_i \|W_{:,i}\|_2^2, \end{aligned} \tag{27}$$

with equality when $a = e_{i^*}$, where $i^* = \arg\min_i \|W_{:,i}\|_2$ corresponds to the *smallest* norm column. With the one-hot vector, the cross terms $w_i w_j$ are 0 if $i \neq j$, and the latter inequalities follow automatically. As a result, the *unique* minimizer of the LHS of Eq. (10) is also a minimizer of the RHS (10).

B. Additional Results

We observe that most methods display a logarithmic-shaped curve for image-based metrics over time. Based on this observation, we find that comparing methods very early or very late in training was not very informative. Too early and the methods have not had sufficient time to accumulate errors; too late and the methods will have chosen many of the same views, leading to similar reconstruction quality. In addition, neither are common situations, as the view selection seeks to retrieve the highest fidelity scene representation as efficiently as possible.

Therefore, we quantitatively compare all methods by showing image-based metrics at 10K (Tabs. 3 to 5) and 30K (Tabs. 6 to 8) gradient steps. In addition, we report an Area-Under-Curve (AUC) metric (normalized by the number of gradient steps), the integrated difference between a method and random over time. We denote this metric with a Δ. The AUC metric indicates the time-averaged improvement of the method over random. We observe that CONVERGE consistently outperforms the competing methods regardless of the sweep over time.

Table 3. Image metrics across several view selection metrics for the Tanks and Temples dataset after 10K gradient steps. Two scenes are visualized, while averages are computed across the entire dataset.

Scene	Method	PSNR↑	ΔPSNR	SSIM↑	ΔSSIM	LPIPS↓	ΔLPIPS
Ignatius	Bayes' Rays	16.71	-0.24	0.43	-0.12	0.56	0.24
	FisherRF	18.01	-0.56	0.59	-0.03	0.31	0.02
	Random	18.45	–	0.64	–	0.28	–
	CONVERGE (Ours)	**19.40**	**0.45**	**0.67**	**0.01**	**0.27**	**-0.01**
Train	Bayes' Rays	12.70	-1.85	0.45	-0.07	0.73	0.28
	FisherRF	15.03	-0.77	0.57	-0.04	0.39	0.04
	Random	16.37	–	0.64	–	0.32	–
	CONVERGE (Ours)	**17.00**	**0.65**	**0.66**	**0.02**	**0.30**	**-0.02**
Overall	Bayes' Rays	14.46	-1.86	0.42	-0.13	0.67	0.29
	FisherRF	16.25	-1.36	0.57	-0.05	0.36	0.05
	Random	18.00	–	0.65	–	0.29	–
	CONVERGE (Ours)	**18.34**	**0.28**	**0.66**	**0.01**	**0.28**	**-0.01**

Table 4. Image metrics across several view selection metrics for the custom dataset after 10K gradient steps. Two scenes are visualized, while averages are computed across the entire dataset.

Scene	Method	PSNR↑	ΔPSNR	SSIM↑	ΔSSIM	LPIPS↓	ΔLPIPS
Shiny	Bayes' Rays	13.08	-1.43	0.31	-0.09	0.88	0.29
	FisherRF	13.69	-1.35	0.38	-0.05	0.58	0.07
	Random	15.47	–	0.46	–	0.47	–
	CONVERGE (Ours)	**16.49**	**0.76**	**0.49**	**0.02**	**0.43**	**-0.03**
Space	Bayes' Rays	11.70	-1.74	0.20	-0.13	0.85	0.34
	FisherRF	13.94	-0.55	0.39	-0.03	0.46	0.03
	Random	14.88	–	0.45	–	0.41	–
	CONVERGE (Ours)	**15.66**	**0.78**	**0.48**	**0.03**	**0.36**	**-0.03**
Overall	Bayes' Rays	13.20	-1.62	0.27	-0.09	0.84	0.35
	FisherRF	14.58	-1.02	0.38	-0.05	0.48	0.06
	Random	16.11	–	0.46	–	0.38	–
	CONVERGE (Ours)	**16.94**	**0.68**	**0.50**	**0.02**	**0.35**	**-0.03**

Table 5. Image metrics across several view selection metrics for the MipNeRF360 dataset after 10K gradient steps. Two scenes are visualized, while averages are computed across the entire dataset.

Scene	Method	PSNR↑	ΔPSNR	SSIM↑	ΔSSIM	LPIPS↓	ΔLPIPS
Counter	Bayes' Rays	19.74	-2.51	0.61	-0.15	0.55	0.30
	FisherRF	22.22	-1.66	0.79	-0.04	0.25	0.04
	Random	24.06	–	0.83	–	0.22	–
	CONVERGE (Ours)	**25.24**	**0.71**	**0.86**	**0.02**	**0.18**	**-0.02**
Garden	Bayes' Rays	17.62	-5.32	0.31	-0.31	0.77	0.51
	FisherRF	23.73	-0.95	0.72	-0.04	0.21	0.03
	Random	25.61	–	0.79	–	0.15	–
	CONVERGE (Ours)	**26.34**	**0.46**	**0.81**	**0.01**	**0.15**	**-0.01**
Overall	Bayes' Rays	18.07	-2.94	0.44	-0.17	0.68	0.35
	FisherRF	20.99	-1.40	0.67	-0.04	0.29	0.05
	Random	22.89	–	0.72	–	0.23	–
	CONVERGE (Ours)	**23.36**	**0.29**	**0.72**	**0.00**	**0.22**	**-0.01**

C. Optimization in Continuous Space and Transmittance-based Metrics

Additionally, we investigate the viability of transmittance-based view metrics and the use of CONVERGE in a continuous space optimization scheme. To test, we compare Equation (17) and CONVERGE against an extension (Gradient), which initializes around the best neighbor (root) and then performs 50 gradient descent steps into the camera pose to minimize coverage, showcasing free-view selection

Table 6. Image metrics across several view selection metrics for the Tanks and Temples dataset after 30K gradient steps. Two scenes are visualized, while averages are computed across the entire dataset.

Scene	Method	PSNR↑	ΔPSNR	SSIM↑	ΔSSIM	LPIPS↓	ΔLPIPS
Ignatius	Bayes' Rays	17.10	-2.06	0.46	-0.20	0.45	0.25
	FisherRF	20.95	-0.28	0.71	-0.03	0.22	0.02
	Random	20.89	–	0.72	–	0.21	–
	CONVERGE (Ours)	**21.53**	**0.68**	**0.73**	**0.02**	**0.20**	**-0.01**
Train	Bayes' Rays	13.04	-4.21	0.46	-0.20	0.70	0.41
	FisherRF	18.93	-1.34	0.72	-0.06	0.25	0.07
	Random	19.97	–	0.76	–	0.19	–
	CONVERGE (Ours)	**20.30**	**0.60**	**0.77**	**0.01**	**0.18**	**-0.02**
Overall	Bayes' Rays	15.03	-3.91	0.43	-0.23	0.62	0.37
	FisherRF	20.26	-1.32	0.71	-0.06	0.23	0.05
	Random	21.12	–	0.75	–	0.20	–
	CONVERGE (Ours)	**21.52**	**0.41**	**0.76**	**0.01**	**0.19**	**-0.01**

Table 7. Image metrics across several view selection metrics for the custom dataset after 30K gradient steps. Two scenes are visualized, while averages are computed across the entire dataset.

Scene	Method	PSNR↑	ΔPSNR	SSIM↑	ΔSSIM	LPIPS↓	ΔLPIPS
Shiny	Bayes' Rays	13.38	-3.08	0.30	-0.17	0.86	0.42
	FisherRF	16.97	-1.59	0.52	-0.06	0.41	0.09
	Random	18.27	–	0.56	–	0.33	–
	CONVERGE (Ours)	**18.41**	**0.59**	**0.58**	**0.02**	**0.31**	**-0.03**
Space	Bayes' Rays	11.76	-3.85	0.19	-0.27	0.83	0.46
	FisherRF	18.16	-0.28	0.58	-0.02	0.29	0.03
	Random	17.85	–	0.58	–	0.27	–
	CONVERGE (Ours)	**18.24**	**0.69**	**0.59**	**0.02**	**0.25**	**-0.03**
Overall	Bayes' Rays	13.40	-3.42	0.27	-0.21	0.82	0.46
	FisherRF	17.94	-1.14	0.54	-0.06	0.33	0.08
	Random	18.71	–	0.58	–	0.27	–
	CONVERGE (Ours)	**19.05**	**0.60**	**0.60**	**0.02**	**0.25**	**-0.02**

Table 8. Image metrics across several view selection metrics for the MipNeRF360 dataset after 30K gradient steps. Two scenes are visualized, while averages are computed across the entire dataset.

Scene	Method	PSNR↑	ΔPSNR	SSIM↑	ΔSSIM	LPIPS↓	ΔLPIPS
Counter	Bayes' Rays	20.39	-4.90	0.62	-0.22	0.48	0.32
	FisherRF	26.35	-1.73	0.87	-0.04	0.17	0.04
	Random	27.51	–	0.89	–	0.15	–
	CONVERGE (Ours)	**27.96**	**0.79**	**0.90**	**0.02**	**0.14**	**-0.02**
Garden	Bayes' Rays	19.28	-7.33	0.33	-0.44	0.75	0.59
	FisherRF	27.77	-1.01	0.85	-0.04	0.12	0.03
	Random	27.85	0.00	0.85	0.00	0.11	0.00
	CONVERGE (Ours)	**28.06**	**0.51**	**0.86**	**0.01**	**0.10**	**-0.01**
Overall	Bayes' Rays	18.81	-4.92	0.45	-0.25	0.63	0.41
	FisherRF	24.92	-1.35	0.78	-0.03	0.19	0.04
	Random	25.57	0.00	0.78	0.00	0.17	0.00
	CONVERGE (Ours)	**25.78**	**0.35**	**0.78**	**0.00**	**0.17**	**-0.01**

and continuous optimization. Since our datasets are finite, we instead take renders from a trained Splatfacto model at these optimized poses and add them to the training set, in addition to the root's frame to stabilize training.

CONVERGE is superior to Equation (17), suggesting that designing the view metric to accommodate transmittance effects when they can be noisy and transient degrades optimality of the chosen viewpoint. CONVERGE outperforms Gradient because Gradient's training images are partially derived from a 3DGS rather than reality. Note that not

Table 9. Image metrics across different view-selection methods at 30K steps, averaged over all scenes.

Setting	Method	PSNR ↑	SSIM ↑	LPIPS ↓
Embodied + Sparse	Equation (17)	21.95	0.695	0.251
	Gradient	21.77	0.701	0.275
	Coverage	**22.39**	**0.707**	**0.244**

including the root frames (which came from ground-truth images) in the training set destabilizes training, which suggests that training a 3DGS using auxiliary 3DGS renders can lead to negative feedback. We believe with access to a simulator or execution on hardware, Gradient might slightly outperform CONVERGE.