An efficient solver for designing optimal sampling schemes

Filip Elvander*, Johan Swärd, and Andreas Jakobsson†
*Dept. of Electrical Engineering (ESAT-STADIUS), KU Leuven, Belgium
†Centre for Mathematical Sciences, Lund University, Sweden

***Abstract*—In this short paper, we describe an efficient numerical solver for the optimal sampling problem considered in *Designing Sampling Schemes for Multi-Dimensional Data* [1]. An implementation may be found on *https://www.maths.lu.se/staff/andreas-jakobsson/publications/*.**

***Index Terms*—Optimal sampling, convex optimization.**

I. Problem statement

For a background to the optimal sampling problem, see [1]. Consider the signal model

$$y(t) \sim p(\cdot; t, \theta)$$

where p is a probability density function parametrized by the sampling parameter $t \in \mathbb{R}^s$ and the parameter vector $\theta \in \mathbb{R}^P$. Here, θ is the parameter of interest to be estimated. Assume that we get to choose to sample y at K out of N potential samples t_n, $n = 1, \ldots, N$. We then want to solve

$$\begin{aligned} \underset{w \in \mathcal{W}, \mu \in \mathbb{R}^P}{\text{minimize}} \quad & \sum_{p=1}^{P} \psi_p \mu_p \\ \text{subject to} \quad & \begin{bmatrix} \sum_{n=1}^N w_n F_n(\theta) & e_p \\ e_p^T & \mu \end{bmatrix} \succeq 0 \\ & p = 1, \ldots, P, \end{aligned} \tag{1}$$

where

$$\mathcal{W} = \left\{ w \in \mathbb{R}^N \mid \sum_{n=1}^N w_n = K \ , \ w_n \in [0,1] \right\}$$

is the set of allowed weights, indicating which K samples that are selected, $\mu \in \mathbb{R}^P$ correspond to the Cramér-Rao lower bound for θ, and where $\psi \in \mathbb{R}_+^P$ is a vector of non-negative weights. Also, e_p is the pth canonical basis vector. Herein, we describe how to solve (1) efficiently by considering its dual problem.[1]

II. Dual problem

Consider the Lagrangian relaxation of (1) according to

$$\mathcal{L} = \sum_p \psi_p \mu_p - \sum_{p=1}^P \left\langle G_p, \begin{bmatrix} \sum_{n=1}^N w_n F_n(\theta) & e_p \\ e_p^T & \mu_p \end{bmatrix} \right\rangle,$$

[1]An implementation of the solution algorithm may be found on *https://www.maths.lu.se/staff/andreas-jakobsson/publications/*.

Algorithm 1 Sub-gradient ascent.

Require: Initial guess $\mathcal{G} = \{G_p\}_{p=1}^P$, step size α.
while Not converged **do**
  Find $w \in \underset{w \in \mathcal{W}}{\arg\min} \ -\sum_{n=1}^N w_n \xi_n(\mathcal{G})$.
  for p=1:P **do**
    $\mu_p \leftarrow e_p^T \left(\sum_{n=1}^N w_n F_n(\theta) \right)^{-1} e_p$.
  end for
  for p=1:P **do**
    $G_p \leftarrow \mathcal{P}_{\mathcal{K}_{\psi_p}} \left(G_p + \alpha \nabla q_p(\mathcal{G}) \right)$.
  end for
  $\mathcal{G} \leftarrow \{G_p\}_{p=1}^P$.
end while
return $w \in \underset{w \in \mathcal{W}}{\arg\min} \ -\sum_{n=1}^N w_n \xi_n(\mathcal{G})$.

where G_p, $p = 1, \ldots, P$, are dual variables, i.e., positive semi-definite matrices of dimension $P \times P$. Let G_p be structured according to

$$G_p = \begin{bmatrix} \tilde{G}_p & \gamma_p \\ \gamma_p^T & g_p \end{bmatrix}. \tag{2}$$

For notational convenience, let a dual point be denoted

$$\mathcal{G} = \{G_p\}_{p=1}^P \tag{3}$$

and define

$$\xi_n(\mathcal{G}) = \left\langle F_n(\theta), \sum_{p=1}^P \tilde{G}_p \right\rangle \tag{4}$$

Then, for any w,

$$\inf_{\mu} \mathcal{L} = \begin{cases} -\sum_{n=1}^N w_n \xi_n(\mathcal{G}) - 2\sum_{p=1}^P e_p^T \gamma_p, & \text{if } g_p = \psi_p, \\ -\infty & \text{otherwise.} \end{cases}$$

The infimum with respect to $w \in \mathcal{W}$ is given by setting the K entries corresponding to the K largest values of $\xi_n(\mathcal{G})$ equal to 1 and the rest to zero. Note that the minimizing w is not necessarily unique. Specifically, if the $K+1$:th largest value of $\xi_n(\mathcal{G})$ is strictly smaller than the K:th largest, the minimizing

w is unique. Otherwise, there are infinitely many solutions. Thus, the dual problem is

$$\begin{array}{ll} \underset{\substack{G_p \succeq 0 \\ p=1,\dots,P}}{\text{maximize}} & \inf_{w \in \mathcal{W}} -\sum_{n=1}^{N} w_n \xi_n(\mathcal{G}) - \sum_{p=1}^{P} e_p^T \gamma_p \\ \text{subject to} & g_p = \psi_p \ , \ p = 1, \dots, P. \end{array}$$

Letting $E = e_P e_P^T$, we may express the constraint as

$$\langle G_p, E \rangle = \psi_p \ , \ p = 1, \dots, P.$$

Thus, defining the family of sets parametrized by ϕ,

$$\mathcal{K}_\phi = \{U \mid \langle U, E \rangle = \phi \ , \ U \succeq 0\} \tag{5}$$

and letting

$$\chi_\psi = \left\{\mathcal{G} \mid G_p \in \mathcal{K}_{\psi_p} \ , \ p = 1, \dots, P\right\}$$

we may express the dual problem as

$$\underset{\mathcal{G} \in \chi_\psi}{\text{maximize}} \ q(\mathcal{G}) \tag{6}$$

where the dual objective function is

$$q(\mathcal{G}) = \inf_{w \in \mathcal{W}} -\sum_{n=1}^{N} w_n \xi_n(\mathcal{G}) - \sum_{p=1}^{P} e_p^T \gamma_p. \tag{7}$$

We utilize the ideas from Nedic and Ozdaglar [2] in order to maximize the dual problem (6) using sub-gradient ascent. The algorithm is summarized in Algorithm 1. A short derivation of the step is presented in the following sections.

A. Sub-gradient ascent

For a dual point $\mathcal{G} \in \chi_\psi$, a sub-gradient of q in (7) at $\mathcal{G}$, denoted $\nabla q(\mathcal{G})$, can be decomposed in components $\nabla q_p(\mathcal{G})$, where each component is given by

$$\nabla q_p(\mathcal{G}) = -\begin{bmatrix} \sum_{n=1}^{N} w_n F_n(\theta) & e_p \\ e_P^T & \mu_p \end{bmatrix} \tag{8}$$

where

$$(w, \mu_p) \in \underset{w, \mu_p}{\arg\min} \ \mathcal{L}(\mu, w, \mathcal{G}). \tag{9}$$

As noted earlier, one may retrieve a primal vector w from this set setting the entries of w corresponding to the K largest values of $\{\xi_n(\mathcal{G})\}_{n=1}^{N}$ to 1 and the rest to zero. Noting that any $\mu_p \in \mathbb{R}$ is a member of the minimizing set, one may here choose

$$\mu_p = e_p^T \left(\sum_{n=1}^{N} w_n F_n(\theta)\right)^{-1} e_p, \tag{10}$$

i.e., the μ_p minimizing the primal objective, while still retaining primal feasibility for this choice of w. Then, a dual ascent method guaranteeing that the dual variable $\mathcal{G}$ is feasible may be realized according to

$$G_p \leftarrow \mathcal{P}_{\mathcal{K}_{\psi_p}} \left(G_p + \alpha \nabla q_p(\mathcal{G})\right)$$

for $p = 1, \dots, P$, where $\mathcal{P}_{\mathcal{K}_\phi}$ denotes projection on $\mathcal{K}_\phi$, as defined in (5). How to perform this projection is described in the next section.

B. Projection on PSD cone with constraint

Consider a set of $\mathcal{G} = \{G_m\}_{m=1}^{M}$ of $M \in \mathbb{N}$ symmetric matrices $G_m \in \mathbb{R}^{P \times P}$. Let $\mathcal{C}$ be the set of $P \times P$ positive semidefinite matrices. Here, we are interested in computing the projection on the set

$$\mathcal{K}_\phi = \{U \mid \langle U, E \rangle = \phi \ , \ U \in \mathcal{C}\} \tag{11}$$

for $\phi \in \mathbb{R}_+$ and a symmetric matrix $E \in \mathcal{C}$.

Proposition 1. *The projection on $\mathcal{K}_\phi$, denoted $\mathcal{P}_{\mathcal{K}_\phi}$ is given as*

$$\mathcal{P}_{\mathcal{K}_\phi} : G \mapsto \mathcal{P}_\mathcal{C}(G + \lambda E) \tag{12}$$

where $\mathcal{P}_\mathcal{C}$ denotes projection on $\mathcal{C}$, and where $\lambda \in \mathbb{R}$ is the unique root of the equation

$$\langle \mathcal{P}_\mathcal{C}(G + \lambda E), E \rangle = \phi. \tag{13}$$

Proof. See appendix. □

Remark 1. *It may here be noted that projecting on $\mathcal{C}$ is simply performed by computing an eigenvalue decompostion and setting all negative eigenvalues to zero.*

C. Computational complexity

It may be noted that finding $(w, \mu) \in \underset{w,\mu}{\arg\min} \ \mathcal{L}(\mu, w, \mathcal{G})$ is linear in N and quadratic in P. Performing the gradient step is linear in P, whereas the projection on the dual feasible set is $\mathcal{O}(P^3)$. To see this, it may be noted that in practice, one may solve the equation $\langle \mathcal{P}(G + \lambda E), E \rangle = \phi$ using interval halving techniques, where each evaluation of the right-hand side requires computing one eigenvalue decomposition. The per-iteration complexity for this scheme is thus $\mathcal{O}(P^3)$.

APPENDIX

Proof of Proposition 1. By definition, $U = \mathcal{P}_{\mathcal{K}_\phi}(G)$ solves

$$\underset{U \in \mathcal{K}_\phi}{\text{minimize}} \ \frac{1}{2} \|U - G\|_F^2 , \tag{14}$$

where $\|\cdot\|_F$ is the Frobenius norm. To arrive at a dual formulation, consider the Lagrangian

$$\tilde{\mathcal{L}} = \frac{1}{2} \|U - G\|_F^2 - \lambda \left(\langle U, E \rangle - \phi \right) - \langle \Lambda, U \rangle,$$

with dual variables $\Lambda \in \mathcal{C}$ and $\lambda \in \mathbb{R}$. We may complete the square according to

$$\begin{aligned} &\frac{1}{2} \|U - G\|_F^2 - \langle \lambda E + \Lambda, U \rangle \\ &= \frac{1}{2} \|U - (G + \lambda E + \Lambda)\|_F^2 - \frac{1}{2} \|G + \lambda E + \Lambda\|_F^2 - \frac{1}{2} \|G\|_F^2 . \end{aligned}$$

Then, the infimum of $\tilde{\mathcal{L}}$ with respect to U is given by

$$\inf_U \tilde{\mathcal{L}} = -\frac{1}{2} \|G + \lambda E + \Lambda\|_F^2 + \|G\|_F^2 + \phi\lambda,$$

which is attained for $U = G + \lambda E + \Lambda$. Consider the dual function

$$r(\Lambda, \lambda) = -\frac{1}{2} \|G + \lambda E + \Lambda\|_F^2 + \phi\lambda. \tag{15}$$

For each λ, this is maximized with respect to $\Lambda \in \mathcal{C}$ by

$$\Lambda = \mathcal{P}_{\mathcal{C}} \left(-(G + \lambda E) \right), \tag{16}$$

i.e., Λ is constructed from the negative part of the eigendecomposition of $G + \lambda E$, with flipped sign. Using

$$G + \lambda E = \mathcal{P}_{\mathcal{C}} \left(G + \lambda E \right) + \mathcal{P}_{\mathcal{C}} \left(-(G + \lambda E) \right), \tag{17}$$

this yields

$$\tilde{r}(\lambda) = \sup_{\Lambda} r = -\frac{1}{2} \|\mathcal{P}_{\mathcal{C}} \left(G + \lambda E \right)\|_F^2 + \phi\lambda. \tag{18}$$

This one-dimensional criterion may then be maximized with respect to $\lambda \in \mathbb{R}$. However, as we from the analysis obtain $U = \mathcal{P}_{\mathcal{C}} \left(G + \lambda E \right)$, we may utilize the primal feasibility condition $\langle U, E \rangle = \phi$. Specifically, one may seek the roots of

$$f(\lambda) = \langle \mathcal{P}_{\mathcal{C}}(G + \lambda E), E \rangle - \phi. \tag{19}$$

As $E \in \mathcal{C}$, f is a continuous, monotone increasing function, and f thus has a unique zero. $\square$

REFERENCES

[1] J. Swärd, F. Elvander, and A. Jakobsson, "Designing Sampling Schemes for Multi-Dimensional Data," *Signal Processing*, vol. 150, pp. 1–10, 9 2018.

[2] A. Nedic and A. Ozdaglar, "Approximate Primal Solutions and Rate Analysis for Dual Subgradient Methods," *SIAM J. Optim.*, vol. 19, no. 4, pp. 1757–1780, 2009.