Note on a Randomized Kernel Ridge Regression and Constrained Quadratic Problem

Jonas Bresch*
Technische Universität Berlin
based on Inspiration at the CU Boulder,
during a research visit at the group of Stephen Becker

Randomization in numerical computations gain in the recent years due to several reasons. Firstly, the computational complexity can be reduced incrementally, secondly, in higher dimensions or due to black-box implementations there is a restricted excess to the data, so that normally one only is allowed to use matrix-vector and vector-vector products. Thirdly, deterministic approaches and algorithms also have to deal with machine precession and adding uncertainties does not effect the computational result as much as the variance can be reduced, ideally.

Randomized Kernel Ridge Regression *Kernel ridge regression* (KRR) is a classical regularized learning method that combines the stability of ridge regression with the flexibility of nonlinear feature spaces [7; 8; 16]. Given training data $\{(x_i, y_i)\}_{i=1}^d$, where x_i lies in an input space $\mathcal{X}$ and $y_i \in \mathbb{R}$, KRR seeks a predictor that fits the data while controlling model complexity through an *reproducing kernel Hilbert space* (RKHS) norm penalty [1]. This has variouse applications via statistical learning [8] with machine learning and support vector machines [13; 14] The key idea is to replace explicit finite-dimensional features by a positive semidefinite kernel function

$$\kappa : \mathcal{X} \times \mathcal{X} \to \mathbb{R},$$

meaning that

$$\sum_{i,j=1}^{d} c_i, c_j \kappa(x_i, x_j) \geq 0, \quad \forall x_i, x_j \in \mathcal{X}, c_i \in \mathbb{R}, d \in \mathbb{N}, \tag{1}$$

which implicitly defines a (possibly infinite-dimensional) RKHS. Here, a positive definite kernel would fulfill equylity to zero in (1) if and only if $c_i = 0$ for all $i \in \mathbb{N}$.

A standard formulation of KRR is

$$\min_{f \in \mathcal{H}_\kappa} \frac{1}{d} \sum_{i=1}^{d} \big(y_i - f(x_i)\big)^2 + \lambda \|f\|_{\mathcal{H}_\kappa}^2,$$

*Webpage and Email

24th May 2026

where $\lambda > 0$ is the regularization parameter and $\mathcal{H}_\kappa$ is the RKHS associated with κ. By the representer theorem, the minimizer admits an expansion of the form

$$f(\cdot) = \sum_{j=1}^{d} v_j\, \kappa(x_j, \cdot),$$

for some coefficients $v = (v_1, \ldots, v_d)^\top \in \mathbb{R}^d$. Writing the kernel matrix as

$$K_{ij} \coloneqq \kappa(x_i, x_j), \qquad i, j = 1, \ldots, d,$$

the vector of fitted values is Kv, and the RKHS norm becomes

$$\|f\|^2_{\mathcal{H}_\kappa} = \langle v, Kv\rangle.$$

Hence the optimization problem can be written equivalently as

$$\min_{v\in\mathbb{R}^d} \frac{1}{d}\|y - Kv\|^2 + \lambda\,\langle v, Kv\rangle, \tag{2}$$

which is the finite-dimensional form used in practice. This representation shows that KRR is a linear least-squares problem in the coefficients v, but with a data-dependent geometry induced by the kernel matrix [7]. The choice of kernel determines the class of functions that can be learned, allowing KRR to model nonlinear structure while retaining convexity and a closed-form solution.

KRR is closely related to ridge regression, Gaussian process regression, and smoothing spline methods, and it is widely used because of its simplicity, theoretical guarantees, and computational tractability for moderate sample sizes.

Solving (2) directly is theoretically possible by evaluating the derivative with respect to v given as

$$\begin{aligned} \nabla_v\big(\tfrac{1}{d}\|y - Kv\|^2 + \lambda\,\langle v, Kv\rangle\big) &= \tfrac{1}{d}K(y - Kv) + 2\lambda Kv = 0 \\ \Leftrightarrow \quad \tfrac{1}{d}Ky &= \big(\tfrac{1}{d}K^2 - 2\lambda K\big)v \\ \Leftrightarrow \quad v &= \big(K^2 - 2d\lambda K\big)^{-1}Ky. \end{aligned}$$

The latter ansatz is numerical complex and inaccurate due to the inverse of $K^2 - 2d\lambda K$ which also involves the matrix-matrix product of K, we all want to avoid.

In the manner of [3–5] we translates to an optimization problem (2) to $\mathbb{S}^{d-1}\times\mathbb{R} \ni (v,\gamma)$ by

$$\min_{v\in\mathbb{R}^d} \frac{1}{d}\|y - Kv\|_2^2 + \lambda\langle v, Kv\rangle = \min_{\substack{\gamma\in\mathbb{R}\\ v\in\mathbb{S}^{d-1}}} \frac{1}{d}\|y - K(\gamma v)\|_2^2 + \lambda\langle \gamma v, K(\gamma v)\rangle$$

where we iteratively aim to solve

$$\operatorname*{arg\,min}_{\tau\in\mathbb{R},\gamma\in\mathbb{R}} \tfrac{1}{d}\|y - \gamma K\tfrac{v+\tau x}{(1+\tau^2)^{1/2}}\|^2 + \lambda\gamma^2\tfrac{\langle v+\tau x, K(v+\tau x)\rangle}{1+\tau^2}, \qquad x \sim \mathcal{U}(\mathbb{S}^{d-1}\cap\mathbb{T}_v). \tag{3}$$

The second part is well-known from [4; 5]. We write out the objective from (3) as

$$\begin{aligned} &\tfrac{1}{d}\|y - \gamma K\tfrac{v+\tau x}{(1+\tau^2)^{1/2}}\|^2 + \lambda\gamma^2\tfrac{\langle v+\tau x, K(v+\tau x)\rangle}{1+\tau^2} \eqqcolon \mathcal{F}(\tau,\gamma) \\ &= \|y\|_2^2 - 2\tfrac{\gamma}{\sqrt{1+\tau^2}}\,\langle y, K(v+\tau x)\rangle + \tfrac{\gamma^2}{1+\tau^2}\|K(v+\tau x)\|_2^2 + \tfrac{\lambda\gamma^2}{1+\tau^2}\,\langle v+\tau x, K(v+\tau x)\rangle \end{aligned}$$

$$= \|y\|_2^2 - 2\tfrac{\gamma}{\sqrt{1+\tau^2}} \langle y, K(v+\tau x)\rangle + \tfrac{\gamma^2}{1+\tau^2}\left(\|K(v+\tau x)\|_2^2 + \lambda \langle v+\tau x, K(v+\tau x)\rangle\right), \quad (4)$$

which is quadratic in γ. Now, differentiating $\mathcal{F}(\tau,\gamma)$, cf. (4), with respect to γ yields

$$\begin{aligned}
& 0 = \partial_\gamma \mathcal{F}(\tau,\gamma) \\
\Leftrightarrow \quad & 0 = -\tfrac{1}{\sqrt{1+\tau^2}} \langle y, K(v+\tau x)\rangle + \tfrac{\gamma}{1+\tau^2}\|K(v+\tau x)\|_2^2 \\
& \quad + \tfrac{\lambda\gamma}{1+\tau^2}\langle v+\tau x, K(v+\tau x)\rangle \\
\Leftrightarrow \quad & \tfrac{1}{\sqrt{1+\tau^2}} \langle y, K(v+\tau x)\rangle = \tfrac{\gamma}{1+\tau^2}\left(\|K(v+\tau x)\|_2^2 + \lambda\langle v+\tau x, K(v+\tau x)\rangle\right) \\
\Leftrightarrow \quad & \langle y, K(v+\tau x)\rangle = \tfrac{\gamma}{\sqrt{1+\tau^2}}\langle (K+\lambda I_d)(v+\tau x), K(v+\tau x)\rangle \\
\Leftrightarrow \quad & \langle y, K(v+\tau x)\rangle = \tfrac{\gamma}{\sqrt{1+\tau^2}}\left\langle v+\tau x, (K^2+\lambda K)(v+\tau x)\right\rangle \\
\Leftrightarrow \quad & \gamma = \tfrac{\sqrt{1+\tau^2}\langle y, K(v+\tau x)\rangle}{\langle v+\tau x, (K^2+\lambda K)(v+\tau x)\rangle} =: \gamma^*(\tau). \qquad (5)
\end{aligned}$$

Notably, the condition

$$\begin{aligned}
& 0 \neq \left\langle v+\tau x, (K^2+\lambda K)(v+\tau x)\right\rangle \\
\Leftrightarrow \quad & \left\langle v+\tau x, K^2(v+\tau x)\right\rangle \neq -\lambda\langle v+\tau x, K(v+\tau x)\rangle \\
\Leftrightarrow \quad & \lambda \neq -\tfrac{\langle v+\tau x, K^2(v+\tau x)\rangle}{\langle v+\tau x, K(v+\tau x)\rangle},
\end{aligned}$$

such that the optimal γ depending on τ exists, is equivalent to the condition that the next iterate $v+\tau x$ for an optimal τ is not an generalized eigenpair of (K^2, K) with generalized eigenvalue $-\lambda < 0$.

Now, replacing γ with the optimal curve $\gamma^*(\tau)$ in $\mathcal{F}(\tau,\gamma)$ yields

$$\mathcal{F}(\tau,\gamma^*(\tau)) = \|y\|_2^2 - \tfrac{\langle y, K(v+\tau x)\rangle^2}{\langle v+\tau x, (K^2+\lambda K)(v+\tau x)\rangle}, \qquad (6)$$

which can be optimized with respect to τ. Therefore, notice that since K is symmetric also K^2 and hence $K^2+\lambda K$ is symmetric, and it is of the form

$$\mathcal{F}(\tau,\gamma^*(\tau)) = \|y\|_2^2 - \tfrac{(\alpha+\tau\beta)^2}{\mathfrak{a}+2\tau\mathfrak{b}+\tau^2\mathfrak{c}},$$

with

$$\begin{aligned}
&\alpha \coloneqq \langle y, Kv\rangle, \quad \beta \coloneqq \langle y, Kx\rangle, \\
&\mathfrak{a} \coloneqq \langle v, (K^2+\lambda K)v\rangle, \quad \mathfrak{b} \coloneqq \langle v, (K^2+\lambda K)x\rangle, \quad \mathfrak{c} \coloneqq \langle x, (K^2+\lambda K)x\rangle.
\end{aligned} \qquad (7)$$

The optimal criteria of the latter are then

$$\begin{aligned}
& 0 = \partial_\tau\left(\tfrac{(\alpha+\tau\beta)^2}{\mathfrak{a}+2\tau\mathfrak{b}+\tau^2\mathfrak{c}}\right) \\
\Leftrightarrow \quad & 0 = \tfrac{2(\alpha+\tau\beta)\beta(\mathfrak{a}+2\tau\mathfrak{b}+\tau^2\mathfrak{c}) - 2(\alpha+\tau\beta)^2(\mathfrak{b}+\tau\mathfrak{c})}{(\mathfrak{a}+2\tau\mathfrak{b}+\tau^2\mathfrak{c})^2} \\
\Leftrightarrow \quad & 0 = (\alpha+\tau\beta)\beta(\mathfrak{a}+2\tau\mathfrak{b}+\tau^2\mathfrak{c}) - (\alpha+\tau\beta)^2(\mathfrak{b}+\tau\mathfrak{c}) \\
\Leftrightarrow \quad & 0 = \beta(\mathfrak{a}+2\tau\mathfrak{b}+\tau^2\mathfrak{c}) - (\alpha+\tau\beta)(\mathfrak{b}+\tau\mathfrak{c}) \\
\Leftrightarrow \quad & 0 = \beta\mathfrak{a} - \alpha\mathfrak{b} + \tau(\beta\mathfrak{b} - \alpha\mathfrak{c}),
\end{aligned}$$

if and only if $0 \neq \alpha+\tau\beta$. Hence the minimizer is given by

$$\tau^* = \tfrac{\beta\mathfrak{a}-\alpha\mathfrak{b}}{\beta\mathfrak{b}-\alpha\mathfrak{c}} \qquad (8)$$

if, additionally, holds $0 \neq \beta\mathfrak{b} - \alpha\mathfrak{c}$. By construction, it should now holds that

$$\|v + \tau^* x\| = \sqrt{1+\tau^2} = \gamma^*(\tau^*). \tag{9}$$

We finally conclude, that

$$\operatorname*{arg\,min}_{\tau\in\mathbb{R}} \frac{1}{d}\|y - K(v+\tau x)\|_2^2 + \lambda\langle(v+\tau x), K(v+\tau x)\rangle = \frac{\langle y,Kx\rangle\langle v,(K^2+\lambda K)v\rangle - \langle y,Kv\rangle\langle v,(K^2+\lambda K)x\rangle}{\langle y,Kx\rangle\langle v,(K^2+\lambda K)x\rangle - \langle y,Kv\rangle\langle x,(K^2+\lambda K)x\rangle}$$

is sufficient for the iterative solution.

Notably, we observe that

$$\begin{aligned}
\mathbb{E}_{x|v}[\langle y, Kx\rangle\langle v, (K^2+\lambda K)x\rangle] &= \mathbb{E}_{x|v}[y^{\mathrm{T}} K x x^{\mathrm{T}} (K^2+\lambda K)v] \\
&= y^{\mathrm{T}} K \mathbb{E}_{x|v}[x x^{\mathrm{T}}](K^2+\lambda K)v \\
&= \tfrac{1}{d-1} y^{\mathrm{T}} K (I_d - v v^{\mathrm{T}})(K^2+\lambda K)v \\
&= \tfrac{1}{d-1}\big(\langle y, K(K^2+\lambda K)v\rangle - \langle y, Kv\rangle\langle v, (K^2+\lambda K)v\rangle\big)
\end{aligned}$$

using that [5]

$$\mathbb{E}_{x|v}[x x^{\mathrm{T}}] = \tfrac{1}{d-1}(I_d - v v^{\mathrm{T}}).$$

Furthermore holds

$$\begin{aligned}
\mathbb{E}_{x|v}[\langle y, Kv\rangle\langle x, (K^2+\lambda K)x\rangle] &= \langle y, Kv\rangle \mathbb{E}_{x|v}[\operatorname{tr}(x^{\mathrm{T}}(K^2+\lambda K)x))] \\
&= \langle y, Kv\rangle \mathbb{E}_{x|v}[\operatorname{tr}((K^2+\lambda K)x x^{\mathrm{T}})] \\
&= \langle y, Kv\rangle \operatorname{tr}((K^2+\lambda K)\mathbb{E}_{x|v}[x x^{\mathrm{T}}]) \\
&= \tfrac{1}{d-1}\langle y, Kv\rangle \operatorname{tr}((K^2+\lambda K)(I_d - v v^{\mathrm{T}})) \\
&= \tfrac{1}{d-1}\langle y, Kv\rangle\big(\operatorname{tr}(K^2+\lambda K) - \operatorname{tr}((K^2+\lambda K)v v^{\mathrm{T}})\big) \\
&= \tfrac{1}{d-1}\langle y, Kv\rangle\big(\operatorname{tr}(K^2+\lambda K) - \operatorname{tr}(\langle v, (K^2+\lambda K)v\rangle)\big) \\
&= \tfrac{1}{d-1}\langle y, Kv\rangle\big(\operatorname{tr}(K^2+\lambda K) - \langle v, (K^2+\lambda K)v\rangle\big)
\end{aligned}$$

such that

$$\begin{aligned}
&\mathbb{E}_{x|v}[\langle y, Kx\rangle\langle v, (K^2+\lambda K)x\rangle - \langle y, Kv\rangle\langle x, (K^2+\lambda K)x\rangle] \\
&\quad = \tfrac{1}{d-1}\big(\langle y, K(K^2+\lambda K)v\rangle - \langle y, Kv\rangle \operatorname{tr}(K^2+\lambda K)\big)
\end{aligned}$$

For the constraints

$$0 \neq \langle y, Kx\rangle\langle v, (K^2+\lambda K)x\rangle - \langle y, Kv\rangle\langle x, (K^2+\lambda K)x\rangle =: \rho(x)$$

we argue via [10: Thm. A.4] and [12]. Therefore, we aim to construct $x \in \mathbb{S}^{d-1} \cap \mathbb{T}_v$ such that $\rho(x) \neq 0$, since

$$N(\rho) = \{w \in \mathbb{S}^{d-1} \cap \mathbb{T}_v \mid \rho(w) = 0\}.$$

is either a null set, or the entire (sub)manifold $\mathbb{S}^{d-1} \cap \mathbb{T}_v \simeq \mathbb{S}^{d-2}$ in $\mathbb{S}^{d-1}$. We determine a case destinction:

1) Let $\langle y, Kv\rangle \neq 0$. Then, we additionally assume that $x \in \operatorname{span}(Ky)^{\perp}$ such that $\langle y, Kx\rangle = \langle Ky, x\rangle = 0$, where $\dim(\operatorname{span}(Ky)^{\perp}) = d-1$ and

$$\dim(\operatorname{span}(Ky)^{\perp} \cap \mathbb{T}_v) \in \{d-2, d-1\},$$

where the latter case is if and only if $v \in \operatorname{span}(Ky)$, such that $\langle y, Kv\rangle = \langle Ky, v\rangle = \pm\|v\|\,\|Ky\| \neq 0$. Now, we would need to have $x \in N(\rho)$ that

$$0 = \langle x, (K^2 + \lambda K)x\rangle = \langle x, K^2 x\rangle + \lambda\langle x, Kx\rangle = \|Kx\|_2^2 + \lambda\langle x, Kx\rangle$$
$$\Leftrightarrow \quad \|Kx\|_2^2 = -\lambda\langle x, Kx\rangle \leq 0.$$

where $\lambda \geq 0$ and $\langle x, Kx\rangle \geq 0$ the positive semidefinite property of the kernel. Hence, if κ is a positive definite kernel, then is K, such that $x = 0$ is necessary, which contradicts $x \in \mathbb{S}^{d-1} \cap \mathbb{T}_v$. Otherwise, we would need that

$$\ker(K) = \operatorname{span}(Ky)^\perp \cap \mathbb{T}_v,$$

since otherwise, we could construct a point in $x \in \ker(K)^c \cap \operatorname{span}(Ky)^\perp \cap \mathbb{S}^{d-1} \cap \mathbb{T}_v$.

2) Let $\langle y, Kv\rangle = 0$. Then we have $\langle y, Kx\rangle \neq 0$ a.s. since otherwise we have $x \in \operatorname{span}(Ky)^\perp$ and hence $v \in \operatorname{span}(Ky)$ which contradicts $\langle y, Kv\rangle = 0$ since then $v = 0$ contradicts the spherical assumption. Similarly it holds

$$(K^2 + \lambda K)v \in \operatorname{span}(v)$$

to have $0 = \langle v, (K^2 + \lambda K)x\rangle$ a.s., i.e. still $x \in N(\rho)$ for (almost any) $x \in \mathbb{S}^{d-1} \cap \mathbb{T}_v$. Hence, v needs to be an eigen vector of $K^2 + \lambda K$.

Bemerkung 1. The null space of K is given by the range of ϕ, where $\phi : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\kappa(x_i, x_j) = \langle \phi(x_i), \phi(x_j)\rangle, \qquad i, j \in \{1, \ldots, d\}.$$

Hence, we assume that $\dim\{\phi(x_1), \ldots, \phi(x_d)\} \geq 3$ such that $\ker(K) \leq d - 3$ and hence $\ker(K) \neq \operatorname{span}(Ky)^\perp \cap \mathbb{T}_v$. ○

Proposition 1. *Let $v \in \mathbb{S}^{d-1}$ no eigenvector of $K^2 + \lambda K$ and $\dim\{\phi(x_1), \ldots, \phi(x_d)\} \geq 3$. Then $\beta\mathfrak{b} - \alpha\mathfrak{c} \neq 0$ a.s.*

Proof. See the argumentation above, similar to [5: Lem. 4.3]. □

Theorem 1. *If $\dim(\phi(x_1), \ldots, \phi(x_d)) \geq 3$ and $\dim(E_\eta(K^2 + \lambda K)) \leq d - 2$ and let $v^0 \sim \mathcal{U}(\mathbb{S}^{d-1})$ and $x^k \sim \mathcal{U}(\mathbb{S}^{d-1} \cap v^k)$ and set $v^{k+1} := \frac{v^k + \tau_k x^k}{(1+\tau_k^2)^{1/2}}$ by τ_k given in (8) with parameters as in (7) then*

1. *v^k is not an eigen vector of $K^2 + \lambda K$ for all $k \in \mathbb{N}$ a.s.,*
2. *$\gamma_k v^k \to \arg\min_{v \in \mathbb{R}^d} \frac{1}{d}\|y - Kv\|_2^2 + \lambda\langle v, Kv\rangle$ where γ_k as in (5).*

Proof. Conclude as in [5: Thm. 5.6] from Prop. 1 via the dimension argument as in [5: Lem. 4.4] and [5: Thm. 4.6]. □

Randomized Constrained Square Problem The optimization problem

$$\min_{v \in \mathbb{R}^d,\ v \geq 0} \langle v, Av\rangle + \langle y, v\rangle, \qquad A \succ 0,\ y \in \mathbb{R}^d, \tag{10}$$

is a simple but fundamental convex quadratic program over the nonnegative orthant [9; 11]. Any (additional) regularization parameter can be absorbed in y. Since the Hessian of the objective is $A + A^*$, where it holds

$$\langle v, Av\rangle = \tfrac{1}{2}\langle v, Av\rangle + \tfrac{1}{2}\langle v, Av\rangle = \tfrac{1}{2}\langle v, Av\rangle + \tfrac{1}{2}\langle A^* v, v\rangle = \Big\langle v, \underbrace{\tfrac{1}{2}(A + A^*)}_{=:A^{\mathrm{H}}} v\Big\rangle$$

and, hence, it is

$$A \succ 0 \quad \Leftrightarrow \quad A^{\mathrm{H}} \succ 0,$$

positive definiteness of A implies strict convexity and therefore a unique minimizer. Problems of this type appear whenever one wants to balance a quadratic energy with a positivity constraint, a structure that is closely related to nonnegative least squares, projected gradient methods, and constrained quadratic programming [2; 6; 11].

Notably, if $y \geq 0$, then $v = 0$ is a minimizer and the minimum is zero, since

$$\langle v, Av \rangle > 0 \quad \forall v \in \mathbb{R}^d, \quad \langle v, y \rangle \geq 0, \quad v \geq 0.$$

From a computational viewpoint, the problem is attractive because it combines a quadratic objective with a very simple feasible set. This makes it a natural test case for active-set, projected-gradient, and operator-splitting methods, all of which are standard tools for linearly constrained convex optimization [2; 6; 15]. Variants of this model also arise in signal processing, inverse problems, and model predictive control, where nonnegativity can encode physical feasibility or interpretability.

Theoretically, (10) is solvable by viewing the Lagrangian. The optimality conditions are given as

$$\nabla_v\big(\langle v, Av \rangle + \langle y, v \rangle\big) = 2A^{\mathrm{H}}v + y = 0 \quad \Leftrightarrow \quad v = \tfrac{1}{2}(A^{\mathrm{H}})^{-1}y$$

and hence, by the nonegativity constraints, it holds

$$v^* = \tfrac{1}{2}((A^{\mathrm{H}})^{-1}y)_+, \quad (\cdot)_+ \coloneqq \max\{0, \cdot\}, \quad \text{componentwise defined.}$$

By assumption A^{H} is invertible but might take numerically a huge effort to build, which we want to avoid.

In the randomized setting, one typically studies how randomization in the data, the matrix A, or the initialization affects the geometry of the feasible region and the behavior of the optimizer [9]. The resulting problems remain convex, but randomization can introduce useful probabilistic structure that simplifies analysis and motivates algorithmic design.

Firstly, we translate (10) to an optimization problem on $\mathbb{S}^{d-1} \times \mathbb{R}$

$$\min_{v \in \mathbb{S}^{d-1}, v \geq 0, \gamma \geq 0} \gamma^2 \langle v, Av \rangle + \gamma \langle v, y \rangle$$

where we iteratively aim to solve

$$\begin{aligned} &\operatorname*{arg\,min}_{\tau \in \mathbb{R}, \gamma \in \mathbb{R}} \gamma^2 \Big\langle \tfrac{v+\tau x}{\|v+\tau x\|}, A\tfrac{v+\tau x}{\|v+\tau x\|} \Big\rangle + \gamma \Big\langle \tfrac{v+\tau x}{\|v+\tau x\|}, y \Big\rangle \eqqcolon \mathcal{Q}(\tau, \gamma) \\ &= \operatorname*{arg\,min}_{\tau \in \mathbb{R}, \gamma \in \mathbb{R}} \gamma^2 \tfrac{\langle v, Av \rangle + 2\tau \langle x, Av \rangle + \tau^2 \langle x, Ax \rangle}{1+\tau^2} + \gamma \tfrac{\langle v, y \rangle + \tau \langle x, y \rangle}{(1+\tau^2)^{1/2}} \quad \begin{cases} x \sim \mathcal{U}(\mathbb{S}^{d-1} \cap \mathbb{T}_v), \\ \quad \text{s.t.} \quad v + \tau x \geq 0. \end{cases} \end{aligned}$$

Secondly, we translate the constrains for the updated variable from before which reads as

$$v + \tau x \geq 0 \quad \Leftrightarrow \quad v \geq \tau x \quad \Leftrightarrow \quad \min_{i: x_i < 0} \{-v_i / x_i\} \geq \tau \geq \max_{i: x_i > 0} \{-v_i / x_i\}.$$

Differentiating with respect to γ yields

$$0 = \partial_\gamma \mathcal{Q}(\tau, \gamma)$$

$$\begin{aligned}
&\Leftrightarrow \quad 0 = 2\gamma\Big\langle \tfrac{v+\tau x}{\|v+\tau x\|}, A\tfrac{v+\tau x}{\|v+\tau x\|}\Big\rangle + \Big\langle \tfrac{v+\tau x}{\|v+\tau x\|}, y\Big\rangle \\
&\Leftrightarrow \quad 0 = 2\gamma\big\langle v+\tau x, A(v+\tau x)\big\rangle + \|v+\tau x\|\big\langle v+\tau x, y\big\rangle \\
&\Leftrightarrow \quad \gamma = -\frac{\|v+\tau x\|\big\langle v+\tau x, y\big\rangle}{2\big\langle v+\tau x, A(v+\tau x)\big\rangle} =: \gamma^*(\tau)
\end{aligned} \tag{11}$$

the optimal curve of γ with respect to τ. Hence we optimize the one-dimensional function

$$\mathcal{Q}(\tau,\gamma^*(\tau)) = \tfrac{\langle v+\tau x, y\rangle^2}{4\langle v+\tau x, A(v+\tau x)\rangle} - \tfrac{\langle v+\tau x, y\rangle^2}{2\langle v+\tau x, A(v+\tau x)\rangle} = -\tfrac{\langle v+\tau x, y\rangle^2}{4\langle v+\tau x, A(v+\tau x)\rangle}$$

Hence, we have

$$\arg\min_{\tau}\ \mathcal{Q}(\tau,\gamma^*(\tau)) = \arg\max_{\tau} \tfrac{\langle v+\tau x, y\rangle^2}{\langle v+\tau x, A(v+\tau x)\rangle} = \arg\max_{\tau} \tfrac{(\alpha+\tau\beta)^2}{\mathfrak{a}+2\tau\mathfrak{b}+\tau^2\mathfrak{c}},$$

in the form as in (6) where we now have

$$\begin{aligned}
&\alpha := \langle v, y\rangle, \quad \beta := \langle x, y\rangle, \\
&\mathfrak{a} := \langle v, Av\rangle, \quad \mathfrak{b} := \langle v, A^{\mathrm{H}}x\rangle = \tfrac{1}{2}(\langle x, Av\rangle + \langle Ax, v\rangle), \quad \mathfrak{c} := \langle x, Ax\rangle.
\end{aligned} \tag{12}$$

We finally conclude with the same argument as in (9) that

$$\arg\min_{\tau\in\mathbb{R}^d} \langle v+\tau x, A(v+\tau x)\rangle + \langle v+\tau x, y\rangle = \tfrac{\langle x,y\rangle\langle v,Av\rangle - \langle v,y\rangle\langle v,A^{\mathrm{H}}x\rangle}{\langle x,y\rangle\langle v,A^{\mathrm{H}}x\rangle - \langle v,y\rangle\langle x,Ax\rangle}.$$

Similarly as before, it holds

$$\begin{aligned}
&\mathbb{E}_{x|v}[\langle x, y\rangle\langle v, A^{\mathrm{H}}x\rangle - \langle v, y\rangle\langle x, Ax\rangle] \\
&= \tfrac{1}{d-1}\big(\langle y, (I_d - vv^{\mathrm{T}})A^{\mathrm{H}}v\rangle - \langle v, y\rangle(\mathrm{tr}(A(I_d - vv^{\mathrm{T}})))\big) \\
&= \tfrac{1}{d-1}\big(\langle y, A^{\mathrm{H}}v\rangle - \langle y, v\rangle \underbrace{\langle v, A^{\mathrm{H}}v\rangle}_{=\langle v,Av\rangle} - \langle v, y\rangle(\mathrm{tr}(A) - \langle v, Av\rangle)\big) \\
&= \tfrac{1}{d-1}\big(\langle y, A^{\mathrm{H}}v\rangle - \langle v, y\rangle\,\mathrm{tr}(A)\big).
\end{aligned}$$

For the constraint that

$$0 \neq \langle x, y\rangle\langle v, A^{\mathrm{H}}x\rangle - \langle v, y\rangle\langle x, Ax\rangle =: \varphi(x)$$

we argue as before via [10: Thm. A.4] and [12] and the nullspace

$$N(\varphi) = \{w \in \mathbb{S}^{d-1} \cap \mathbb{T}_v \mid \varphi(w) = 0\}.$$

1) Let $\langle v, y\rangle = 0$, then we have $\langle x, y\rangle \neq 0$ a.s., since otherwise we would have $y \in \mathrm{span}(v)$, and hence $y = 0$ is necessary. Then, $\langle v, A^{\mathrm{H}}x\rangle = \langle A^{\mathrm{H}}v, x\rangle = 0$ a.s., if and only if v is an eigen vector of A^{H}, cf. [5: Lem. 4.3].

2) Let $\langle v, y\rangle \neq 0$, then we choose additionally $x \in \mathrm{span}(y)^{\perp}$ and notice that

$$\dim(\mathrm{span}(y)^{\perp} \cap T_v) \in \{d-2, d-1\}$$

where the latter case arise if and only if $y \in \mathrm{span}(v)$ such that then we would have $\langle v, y\rangle = \pm\|v\|\,\|y\| \neq 0$ if $y \neq 0$. Hence, we would need for $x \in \mathbb{S}^{d-1} \cap \mathrm{span}(y)^{\perp} \cap T_v$ that $\langle x, Ax\rangle = 0$, which contradicts $A \succ 0$.

Proposition 2. *Let $v \in \mathbb{S}^{d-1}$ no eigenvector of A^{H} and $y \neq 0$. Then $\beta\mathfrak{b} - \alpha\mathfrak{c} \neq 0$ a.s.*

Proof. See the argumentation above, similar to [5: Lem. 4.3]. □

Theorem 2. *If $y \neq 0$ and $\dim(E_\eta(A^{\mathrm{H}})) \leq d-2$ and let $v^0 \sim \mathcal{U}(\mathbb{S}^{d-1})$ and $x^k \sim \mathcal{U}(\mathbb{S}^{d-1} \cap v^k)$ and set $v^{k+1} \coloneqq \frac{v^k + \tau_k x^k}{(1+\tau_k^2)^{1/2}}$ by τ_k given in (8) with parameters in (12) then*

1. *v^k is not an eigen vector of A^{H} for all $k \in \mathbb{N}$ a.s.,*
2. *$\gamma_k v^k \to \arg\min_{v \in \mathbb{R}^d} \langle v, Av \rangle + \langle y, v \rangle$ where γ_k as in (11).*

Proof. Conclude as in [5: Thm. 5.6] from Prop. 2 via the dimension argument as in [5: Lem. 4.4] and [5: Thm. 4.6]. □

References

[1] Aronszajn, N., "Theory of Reproducing Kernels", *Transactions of the American Mathematical Society*, 68/3 (1950), 337–404. doi: 10.1090/S0002-9947-1950-0051437-7.

[2] Bemporad, A., "A Numerically Stable Solver for Positive Semidefinite Quadratic Programs Based on Nonnegative Least Squares", *IEEE Transactions on Automatic Control*, 63/2 (2018), 525–31. doi: 10.1109/TAC.2017.2735938.

[3] Bresch, J., *Generalization of zeroth-order method for quotients of quadratic functions*, arXiv:2604.26913, submitted to J. Optim. Theory Appl. (April, 2026), 2026. doi: 10.48550/arXiv.2604.26913.

[4] Bresch, J., Lorenz, D. A., Schneppe, F., and Winkler, M., *Matrix-free stochastic calculation of operator norms without using adjoints*, arXiv:2410.08297, accepted in SIAM J. Matrix Anal. Appl. (March, 2026), 2024. doi: 10.48550/arXiv.2410.08297.

[5] Bresch, J., Melnyk, O., Schoen, M., and Steidl, G., *Stochastic zeroth-order method for computing the generalized Rayleigh quotient*, arXiv:2512.05520, submitted to Linear Algebra Appl., November, 2025. doi: 10.48550/arXiv.2512.05520.

[6] Calamai, P. H., and Moré, J. J., "Projected Gradient Methods for Linearly Constrained Problems", *Mathematical Programming*, 39/1 (1987), 93–116. doi: 10.1007/BF02592073.

[7] Cheng, N., Doostan, A., and Becker, S., "Stochastic Subspace Descent Accelerated via Bi-fidelity Line Search" (2025), arXiv:2505.00162. doi: 10.48550/arXiv.2505.00162.

[8] Hastie, T., Tibshirani, R., and Friedman, J., *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd edn., New York, NY: Springer, 2009). doi: 10.1007/978-0-387-84858-7.

[9] Jensen, T., Becker, S., and al., et, "A bi-fidelity Proximal Quasi Newton Method for Dense Rigid Body" (2025).

[10] Langrenez, C., Salmon, W., Bièvre, S. D., Thio, J. J., Long, C. K., and Arvidsson-Shukur, D. R. M., *The set of Kirkwood–Dirac positive states is almost always minimal*, arXiv:2405.17557, 2024, https://doi.org/10.48550/arXiv.2405.17557.

[11] Lawson, C. L., and Hanson, R. J., *Solving Least Squares Problems* (Philadelphia, PA: Society for Industrial and Applied Mathematics, 1995). doi: 10.1137/1.9781611971217.

[12] Mityagin, B. S., "The zero set of a real analytic function", *Math. Notes*, 107/3 (2020), 529–30, https://doi.org/10.1134/S0001434620030189.

[13] Sch"olkopf, B., and Smola, A. J., *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond* (Cambridge, MA: MIT Press, 2002). doi: 10.7551/mitpress/4175.001.0001.

[14] Steinwart, I., and Christmann, A., *Support Vector Machines* (Information Science and Statistics; New York, NY: Springer, 2008). doi: 10.1007/978-0-387-77242-4.

[15] Stellato, B., Banjac, G., Goulart, P., Bemporad, A., and Boyd, S., "OSQP: An Operator Splitting Solver for Quadratic Programs", *Mathematical Programming Computation*, 12/4 (2020), 637–72. doi: 10.1007/s12532-020-00179-2.

[16] Wahba, G., *Spline Models for Observational Data* (CBMS-NSF Regional Conference Series in Applied Mathematics; Philadelphia, PA: Society for Industrial and Applied Mathematics, 1990). doi: 10.1137/1.9781611970128.